Title

Information content of spatially distributed ground-based measurements for hydrologic-parameter calibration in mixed rain-snow mountain headwaters

Authors

Avanzi, F
Maurer, T
Glaser, SD
et al.

Publication Date

2020-03-01

DOI

10.1016/j.jhydrol.2019.124478

Peer reviewed

Journal of Hydrology, 2019 https://doi.org/10.1016/j.jhydrol.2019.124478

Information content of spatially distributed ground-based measurements for hydrologic-parameter calibration in mixed rain-snow mountain headwaters

Francesco Avanzi[a,*], Tessa Maurer[a], Steven D. Glaser[a], Roger C. Bales[b,a], Martha H. Conklin[b]

[a] *Department of Civil and Environmental Engineering, University of California, Berkeley, 94720, Berkeley, California, USA*
[b] *Sierra Nevada Research Institute, University of California, Merced, 95343, Merced, California, USA*

Abstract

Parameters in hydrologic models used in mixed rain-snow regions are often uncertain to calibrate and overfitted on streamflow. To contribute addressing these challenges, we used an algorithm that assesses modeling performances through time (Dynamic Identifiability Analysis) to quantify the information content of spatially distributed ground-based measurements for identifying optimal parameter values in the Precipitation Runoff Modeling System (PRMS) model. Including spatially distributed ground-based measurements in Identifiability Analysis allowed us to unambiguously estimate more parameter values than only using streamflow (seven parameters instead of two out of a pool of thirty-three). Peaks in information gain were obtained when using dew-point temperature to identify precipitation phase-partitioning parameters. Multi-attribute identifiability analysis also yielded optimal parameter values that were temporally less variable than those estimated using streamflow alone. Overall, identifying parameter values using ground-based measurements improved the simulation of key drivers of the surface-water budget, such as air temperature and precipitation-phase partitioning. However, parameters simulating surface-to-subsurface mass fluxes like snow accumulation and melt or evapotranspiration were poorly identified by any attribute and so emerged as key sources of predictive uncertainty for this distributed-parameter hydrologic model. This work demonstrates the value of expanded ground-based measurements for identifying parameters in distributed-parameter hydrologic models and so diagnosing their conceptual uncertainty across the water budget.

Keywords: Hydrologic model, Identifiability, Feather River, Parameter calibration, PRMS, Snowmelt.

1. Introduction

In snow-dominated regions like the western United States, winter precipitation and snow-cover accumulation have a dominant impact on water supply (Serreze et al., 1999; Barnett et al., 2005). In such contexts, both hydropower and water providers forecast summer water volume starting from real-time precipitation and streamflow data (Garen, 1992; Georgakakos et al., 2004; Pagano et al., 2004). In California, the state Department of Water Resources (DWR) relies on monthly multiple regressions between seasonal unimpaired flow and water-year precipitation, runoff, and snow water equivalent (SWE, see Rosenberg et al., 2011). Pacific Gas & Electric (PG&E), the largest energy company in northern California, uses similar statistical tools to predict incoming water to reservoirs throughout the southern Cascades and the Sierra Nevada (Freeman, 1999; Koczot et al., 2004).

DWR forecasts tend to perform better in basins where snow-cover accumulation is a major driver of seasonal runoff (Harrison & Bales, 2016). In California as in many other parts of the world, climate change will continue to reduce snow accumulation and thus storage for summer water supplies (Mote, 2003; Knowles et al., 2006; Hatchett et al., 2017; Mote et al., 2018). Coupled with greater evapotranspiration (Goulden & Bales, 2014; Roche et al., 2018), a warming climate affects the timing of seasonal runoff (Musselmann et al., 2017) and the overall water balance (Barnhart et al., 2016; Bales et al., 2018). These shifts, and the inherent variability

*Corresponding author
Email address: francesco.avanzi@polimi.it (Francesco Avanzi)

of climatic conditions, make regression-based approaches vulnerable to unobserved scenarios, with implications for water-resources management and flood control (Freeman, 2015; Harrison & Bales, 2016). In an attempt to make forecasts more robust, mass- and energy-balance hydrologic models have been increasingly deployed in forecasting operations (Koczot et al., 2004; Devineni et al., 2008; Li et al., 2009; Pagano et al., 2014). These models are also used to extend forecast lead times to short-term water supply, in addition to seasonal water volumes.

While all hydrologic models are conceptual to some extent (Wagener et al., 2003), existing approaches can be broadly classified into lumped or distributed (Beven, 2012). At the intersection between fully lumped and fully distributed models, there exist several models of intermediate complexity that respond to physical principles, fragment a basin into smaller Hydrologic Response Units (HRUs), and distribute parameter values to capture the spatial heterogeneity of hydrologic processes (Bartolini et al., 2011; Schaefli et al., 2014; Wi et al., 2015; Markstrom et al., 2016; Bongio et al., 2016). These distributed-parameter models seek an important trade-off across simulation speed, physical realism, and degrees of parametric freedom and are therefore a valid option for operational forecasting (Pagano et al., 2014). For instance, both DWR and PG&E have partnered with the U.S. Geological Survey to implement the Precipitation Runoff Modeling System (PRMS, see Markstrom et al., 2015), a deterministic, distributed-parameter model including tens of calibration parameters (35 out of 108 in Markstrom et al., 2016).

Albeit computationally efficient, distributed-parameter models pose additional calibration challenges, since single-objective strategies may misrepresent internal processes within the water budget while preserving acceptable results in simulating streamflow (Rajib et al., 2016). This overfitted approach increases the overall predictive uncertainty because it skews parameter calibration toward streamflow fit rather than toward maximizing performance for the process that these parameters are supposed to represent (Wagener & Gupta, 2005). Calibration uncertainty and overfitting are common in mountain headwaters across the seasonal rain-snow transition zone, where data are sparse and catchment heterogeneity is high (Saksa et al., 2017). Some recurring sources of predictive uncertainty in these regions include precipitation amount and distribution, precipitation-phase partitioning, peak SWE and snow-melt timing across the landscape, and surface-subsurface exchanges (Bales et al., 2006; Beven & Westerberg, 2011; Avanzi et al., 2014; Harder & Pomeroy, 2014; Harpold et al., 2017; Bales et al., 2018).

Strategies exist to reduce overfitting to streamflow, including multi-objective calibrations (Yapo et al., 1998; Gupta et al., 1998; Hay et al., 2006; Bekele & Nicklow, 2007; Rajib et al., 2016), data assimilation (Mitchell et al., 2004; Andreadis & Lettenmaier, 2006; Clark et al., 2006; Dressler et al., 2006; Andreadis et al., 2007; Massari et al., 2018), transfer functions and regionalization (Parajka et al., 2005; Samaniego et al., 2010), or even manual tuning (Gupta et al., 1999; Boyle et al., 2000). All these approaches rely to some extent on the hypothesis that collecting more measurements will directly translate into improved simulations, for example by reducing calibration uncertainty (Yapo et al., 1998; Hay et al., 2006). This hypothesis has been frequently challenged on both theoretical and practical grounds (see Klemeš, 1977; Kirchner, 2006; Beven & Westerberg, 2011), especially since the correlation between data availability and model performance is often weak (Tang & Lettenmaier, 2010; Pagano et al., 2014).

Remote sensing has demonstrated great potential to constrain model predictions (Silvestro et al., 2015; Nijzink et al., 2018), but much of these data are limited in temporal-spatial resolution, especially in complex terrain (Camici et al., 2018). Ground-based measurements, on the other hand, have a particularly fine temporal resolution that makes them appealing sources of information to detect, for example, the rain-snow transition line and the amount of precipitation across the landscape (Zhang et al., 2017). While ground-based measurements have been used to force or "sanity-check" hydrologic-model predictions since the early stages of operational hydrology (Pagano et al., 2014), their value for improving hydrologic-parameter calibration in mountain regions and thus reducing overall predictive uncertainty has been rarely quantified in a systematic way. Ground-based measurements are sparsely used in model calibration as a result (Bongio et al., 2016; Saksa et al., 2017).

The aim of the present study was to objectively quantify the information content of spatially distributed, mostly ground-based alternative measurements to streamflow for calibrating parameters in distributed-parameter mountain-hydrology models

and so assess how these measurements could be strategically used to address the recurring trade-off between maximizing hydrologic-parameter fit on streamflow or on headwater processes. We used PRMS, a hydrologic model that is operationally deployed in the Sierra Nevada of California, and Dynamic Identifiability Analysis (DYNIA, see Wagener et al., 2003), an algorithm able to detect periods of high information content in data for hydrologic-parameter calibration. The three research questions we set out to answer are: (1) What environmental measurements are the most informative for hydrologic-parameter calibration in mountain, mixed rain-snow headwater basins? (2) In particular, what parameters can be better calibrated using spatially distributed and temporally dense ground-based measurements, versus using streamflow? (3) How does constraining parameter values with alternative measurements to streamflow affect predictive accuracy across the water budget?

2. Material and Methods

Our study area is the headwaters of the Feather River, a major tributary of the Sacramento River in California. Albeit one of the most studied catchments in California from a hydrologic-forecasting standpoint, most research about the Feather has focused on incoming flow to Oroville Dam (see Koczot et al., 2004; Wood & Schaake, 2008; Tang & Lettenmaier, 2010; Huang et al., 2012; Anghileri et al., 2016 and references therein). However, headwaters are a primary source of hydropower (Koczot et al., 2004), which must respond to changes in both seasonal and inter-annual water budgets and market prices (Gaudard et al., 2018).

Data considered in this study included both spatially distributed, ground-based measurements and a reanalysis product for SWE (Section 2.1). In the following, each environmental variable considered in this work will be referred to as an *attribute* (e.g., air temperature, SWE, streamflow); data of a given attribute from a specific location will be referred to as *measurements*. We performed a separate identifiability analysis for each of these attributes and a relatively high number of PRMS parameters (33) covering the entire spectrum of hydrologic processes in headwater catchments (henceforth, multi-attribute identifiability analysis).

Identifiability analysis in hydrology refers to the identification of both an adequate model structure (Gupta et al., 2012) and optimal parameter values (Wagener et al., 2003). DYNIA addresses both issues by stochastically sampling parameter values for a given model and assessing their performance using measurements of a selected output variable (for example, streamflow). By repeating this performance assessment over a moving temporal window, DYNIA identifies (1) periods of high information content to calibrate model parameters; (2) model structural uncertainty; (3) dominant hydrologic processes over specific seasons and years; (4) data anomalies (see again Wagener et al., 2003). In this context, *parameter identification* is conceptually similar to calibration (Wagener et al., 2003). Details about PRMS and about how we performed multi-attribute identifiability with this model are reported in Sections 2.2 and 2.3, respectively.

Results of this multi-attribute identifiability analysis were assessed in view of (1) the most informative attribute to calibrate each of the 33 considered parameters of PRMS; (2) temporal variability of information content and optimal parameter values as estimated through multi-attribute identifiability analysis; (3) changes in PRMS performance obtained by tuning parameters based on multi-attribute identifiability analysis rather than only relying on streamflow (Section 2.4).

2.1. Site description and data

The headwaters of the Feather River drain to Oroville Lake, a reservoir with a total capacity of 4.4 km^3. Oroville Lake is a central component of the State Water Project, a 1100-km network of storage and conveyance infrastructure delivering water to 27 million people and ∼300,000 hectares of irrigated land across California (`https://www.water.ca.gov/Programs/State-Water-Project`, visited July 14, 2019). For simplicity, we refer to the drainage basin upstream of Oroville Lake as the Feather River basin (area of $\sim$ 9300 km^2, Figure 1(a)).

Located at the transition between the Sierra Nevada and the southern Cascade mountain ranges, the bedrock of the Feather River basin is predominantly volcanic in the northern and western portions of the catchment and granitic in the south (Koczot et al., 2004). The climate is Mediterranean, with dry summers and wet winters and springs. Orographic effects dramatically reduce precipitation from the wet, western and northern areas of the basin to the dry, eastern part (see Figure 8 in Koczot et al., 2004). The basin elevation

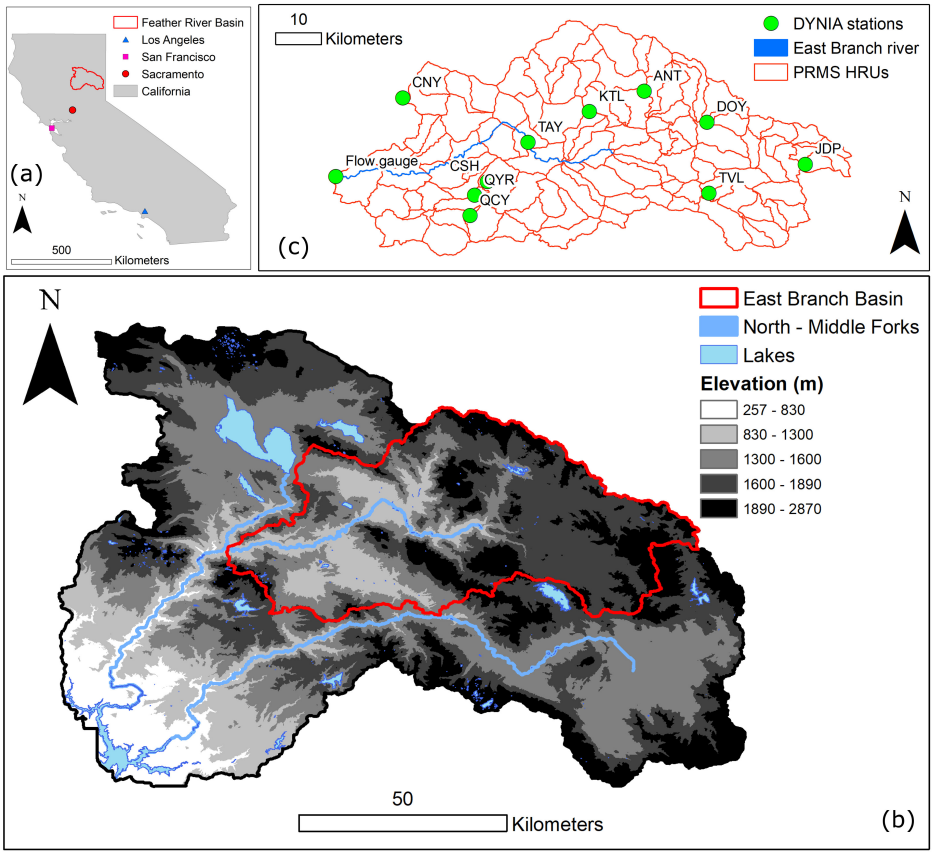

Figure 1: Panel (a): location of the Feather River basin upstream of Oroville Lake; Panel (b): topography and hydrography of the Feather River; Panel (c): location of available ground-based measurements over the East Branch. For context, the Hydrologic Response Units (HRUs) used by PRMS on this Branch are also reported (Koczot et al., 2004).

ranges from 257 m ASL (Oroville Dam) to 2903 m ASL (Lassen Peak), but most of the basin lies between 1000 and 2000 m ASL (see Figure 1(b)) and is thus frequently exposed to mixed rain-snow and rain-on-snow events (Anghileri et al., 2016).

Oroville Lake has four tributaries: the North and the Middle Forks (the longest tributaries, see Figure 1(b)), the South Fork, and the West Branch. With several reservoirs, ten powerhouses, and a total capacity of 734 MW, the North Fork is a major source of hydropower for PG&E. The two main headwater catchments of this fork are the Almanor subbasin (north-west, impounded by Lake Almanor) and the East Branch, a 2600 km^2 rain-shadowed and largely unregulated subbasin (Freeman, 2011), see Figure 1(b). As a mid-latitude catchment straddling the rain-snow transition zone, the East Branch is experiencing decreasing runoff, higher forest density than in the early 20th century, and potentially further increased evapotranspiration due to rising temperatures (Freeman, 2011). For all these reasons, we focused on the East Branch to measure the potential value of multiple, mostly ground-based attributes for hydrologic predictions and operations in mountain regions.

Considered measurements on the East Branch included (1) ground-based weather data from the California Data Exchange Center (CDEC, `https://cdec.water.ca.gov/`, visited July 14, 2019), (2) distributed snow-cover patterns from reanalysis (Margulis et al., 2016), and (3) daily measurements of unimpaired flow at the outlet of the catchment, currently maintained by PG&E (Figure 1(c)). Table 1 summarizes all the ground-based measurements besides unimpaired flow. These measurements were directly downloaded from CDEC at daily resolution or aggregated from sub-daily to daily and were processed using both automatic range checks and visual inspection. We chose a conservative approach and removed any visually sus-

picious data point, that is, any point that significantly deviated from the local main trend (including spikes). No replacement, interpolation, or regionalization was attempted to avoid additional uncertainty in measurements.

As for snow patterns, we used an already validated distributed SWE product (90 m resolution, see Margulis et al., 2016) rather than remotely sensed measurements (such as snow covered area from MODIS) because hydrologic models usually simulate SWE as the primary snow-cover variable of interest. We preferred a reanalysis product over ground-based measurements from snow pillows or manual courses due to the known biases of the first with regard to snow accumulation and melt (see discussions in Johnson & Marks, 2004; Bales et al., 2006) and the coarse temporal resolution of the second (monthly). The chosen reanalysis product covers water years 1985 to 2016.

Measurements of unimpaired flow include water years 1970-2017 and were already quality checked by PG&E. In hydrologic practice in the western United States, a "water year" is the period between October 1 and the following September 30 and is indicated with the calendar year in which it ends.

Because of its importance as a high-elevation site, we expanded the suite of measurements at the CDEC station Kettle Rock (KTL) by installing a wireless-sensor network measuring air temperature, relative humidity, snow depth, and soil-water content and temperature at 12 locations within $\sim$1 km^2 around the preexisting snow pillow. Soil-water content is measured at two depths (25 and 50 cm) and was used to compute depth-integrated soil-water storage across the first meter of soil. We assumed the two readings to be representative of the first 30 and the deeper 70 cm of soil, respectively. Measurement from the 12 locations were averaged to obtain a single time series for each variable (see Table 1). This station is part of a larger deployment called the Feather River Hydrologic Observatory (see deployment details in Malek et al., 2017; Avanzi et al., 2018 – data available at frho.us, visited July 14, 2019).

2.2. The Feather-River PRMS model

PRMS was implemented on the Feather River (Figure 1(b)) in the late 1990s - early 2000s through a collaboration of the US Geological Survey, DWR, and PG&E, (Koczot et al., 2004). While PRMS is now available in version 5 (May 2019), the version available on the river at the time of this research was version 2 (see Leavesley et al., 1983; Koczot et al., 2004).

This model is described and validated in Koczot et al. (2004), so we limit ourselves to a brief overview on its implementation on the East Branch, our main study area. The model divides the East Branch in 111 HRUs, ranging from 4.4 to 54.8 km^2, and then solves mass and energy conservation at the centroid of each HRU (Figure 1(c) and Koczot et al., 2004). The HRU-delineation process started from watershed boundaries of the California State Water Resources Control Board, which were further refined by prioritizing drainage divides and hydrography over other catchment characteristics like canopy coverage and soil properties (Koczot et al., 2004). To better compare local measurements of snow depth and soil-water content with model predictions, we introduced an additional, small HRU (1 km^2) corresponding to station KTL (see Section 2.1). Parameter values at this HRU were either assumed to be the same as those of the original HRU of KTL, or estimated from topographic - canopy information. The small area of this HRU minimizes the impact of these parameters on streamflow prediction.

Input data for this model include daily maximum and minimum air temperature and daily total precipitation. Available input data start in 1969 and were collected directly by PG&E from several sources, with some data not being available from the online sources in Table 1. Air temperature for the East Branch is inputed at one valley location (QCY-QYR in Table 1), which was chosen by Koczot et al. (2004) both because of its central location in the basin and the fact that it was the only one to provide real-time, telemetered data when the model was implemented on the river (this is important since PRMS on the Feather is routinely used in water-supply and hydropower forecasting). From this valley location, the model estimates daily maximum and minimum air temperature at the centroid of each HRU using monthly lapse rates (parameters *tmax_lapse* and *tmin_lapse*, see Koczot et al., 2004 for details).

Precipitation for each HRU centroid is obtained by first fitting a regression plane between latitude and longitude (predictors) and daily precipitation as a percent of normal from ten ground-based stations across the Feather river (target). This regression plane is then multiplied by monthly average precipitation from the Parameter elevation Regression on Independent Slopes Model (PRISM).

Table 1: Ground-based measurements considered in this work, their corresponding elevation, and available periods (calendar year of the first and last data points, respectively).

Station CDEC code	Elevation (m ASL)	Air T.	RH	Solar rad.	Snow depth	Soil WC[a]
ANT	1511	2005 -2017				
CNY	1390	2007 - 2017				
CSH	1378	2005 - 2017	2001 - 2017	2001 - 2017		
DOY	1728	2005 - 2017	2001 - 2014	2013 - 2017		
JDP	2076	2005 - 2017	2005 - 2017	2005 - 2017		
KTL	2225	2005 - 2017	2016 - 2017	2016 - 2017	2016 - 2017	2016 - 2017
QCY	1039	1996 - 2017				
QYR	1066	2005 - 2017	1995 - 2017	2005 - 2017		
TAY	1078	2007 - 2017	2007 - 2017	2007 - 2017		
TVL	1647	2006 - 2017	2006 - 2017	2006 - 2017		

[a]Air T. is air temperature; RH is relative humidity; Solar Rad. is incoming shortwave solar radiation; and Soil VWC is soil-water content (or storage). All measurements are daily averages; Air T. includes daily maximum and minimum. Sources: California Data Exchange Center (CDEC, https://cdec.water.ca.gov/, visited July 14, 2019); Feather River Hydrologic Observatory (frho.us, visited July 14, 2019).

The resulting surface is sampled at the location of HRU centroids to obtain daily precipitation values. This method, called DRAPER, aims to account for both orographic and rain-shadow effects (through PRISM) and day-to-day variability in precipitation patterns (through the ground-based stations), the latter being particularly important in the Feather river due to funneling of moist air and enhancement of precipitation through the canyons of the North Fork (Koczot et al., 2004). PRISM is a frequently used source of distributed precipitation in the western US that has been developed to explicitly account for physiographic features (Daly et al., 2008). PRISM precipitation patterns on the Feather were compared to ground-based gauges by Koczot et al. (2004) and the two data sources agree in terms of orographic and rain-shadow effects. Predictions by DRAPER were not validated directly, given the paucity of precipitation gauges (Koczot et al., 2004). DRAPER is externally run prior to PRMS execution and covers the whole Feather River upstream of Oroville Dam (see details, including code algorithm, at Koczot et al., 2004; Donovan & Koczot, 2019).

The water budget in PRMS is solved by computing precipitation phase partitioning, radiation distribution, canopy interception, snow-cover accumulation and depletion, evapotranspiration, surface runoff, soil-water storage, and groundwater flow. Precipitation-phase partitioning between snow and rain is estimated from daily maximum and minimum temperature at each HRU using two static temperature thresholds, $tmax\_allsnow$ amd $tmax\_allrain$. The model thus calculates a daily fraction of liquid over total precipitation for each HRU on each precipitation day: 0 means that all precipitation at that HRU is snow, 1 means that all precipitation at that HRU is rain. Daily incoming shortwave solar radiation is obtained by relating potential shortwave radiation to daily maximum temperature through a modified degree-day approach (see Markstrom et al., 2015 for all parameters involved). Potential daily evapotranspiration is computed using the Jensen-Haise formulation (the main parameter being jh_coef, a multiplicative term to average daily temperature and shortwave radiation). The amount of precipitation intercepted by canopy depends on vegetation fraction and plant-cover type (parameters being seasonal canopy storage capacities for each phase and seasonal-coverage density); a daily mass balance of this intercepted mass is computed by involving potential evapotranspiration, throughfall, and precipitation.

Snow accumulation and melt is solved by a mass- and energy-balance model, which also accounts for the non-linear relation between SWE and snow-covered area, surface-albedo decay, and sublimation (see Markstrom et al., 2015 for all parameters involved). Surface runoff is modeled using a variable-source-area parametrization in which the surface-runoff generating areas vary with time and with antecedent soil moisture according to three main parameters, $smidx\_coef$, $smidx\_exp$, and $carea\_max$.

The latter represents the maximum possible area contributing to surface runoff, while the first two determine the contributing area. Soil-water storage is split between a surface, finite reservoir accounting for soil moisture (including recharge) and a subsurface, infinite reservoir accounting for interflow (see Markstrom et al., 2015 for all parameters involved). Groundwater flow is modeled using a linear relation with groundwater storage (assumed infinite) via the coefficient $gwflow\_coef$. Besides standard debugging, the most relevant difference between version 2 and 5 is in the subsurface modules (Markstrom et al., 2015), since the most recent versions replaced the finite and infinite reservoirs with three reservoirs accounting for capillary, gravity, and preferential flow.

2.3. Dynamic Identifiability over multiple environmental attributes

The application of DYNIA to multiple attributes (and thus multiple measurement points across the watershed) was based on three main steps (see Figure 2): (1) coupling available measurements with PRMS modeling outputs, (2) running DYNIA for each measurement-model output pair; and (3) comparing the information content of each attribute for identifying optimal values of a given parameter. We implemented this algorithm starting from the SAFE toolbox (v 1.1), a Matlab/Octave software package that implements sensitivity-analysis methods, parameter-sampling tools, DYNIA, and a number of other modeling utilities (see Pianosi et al., 2015 for more details). SAFE is an open-source software available at https://www.safetoolbox.info (visited July 14, 2019).

2.3.1. Step 1

Step 1 was a preliminary task: available measurements were associated to specific PRMS outputs that were assumed to be the modeling counterpart of those observations. The result of this step was a set of measurement-model output pairs: let Z be a vector including all these pairs and Z_i one of these pairs (for example, one Z_i could be composed by measurements of air temperature and simulated air temperature at the same location i).

Measurements of daily maximum and minimum air temperature were compared with corresponding simulated values at select HRU centroids (*tmaxc* and *tminc*, see next paragraph for the selection of centroids). As discussed in Section 2.2, these simulated values were obtained by PRMS based on monthly lapse rates (parameters) and temperature at one input station (input), meaning *tmaxc* and *tminc* are actual state variables in the model. Incoming shortwave solar radiation (henceforth, simply solar radiation), snow depth, SWE, and streamflow were directly compared with the corresponding state variables of PRMS (*swrad*, *pk_depth*, *pkwater_equiv*, and *basin_cms*, respectively). Measurements of soil-water content were compared with simulated soil-water content in the recharge zone (PRMS variable *soil_rechr*). Note that only one water year of measurements was available for snow depth and soil-water content, meaning results for these attributes are of indicative nature.

Measurements of average daily relative humidity were combined with co-located measurements of average daily temperature to estimate dew-point temperature (DPT, see Lawrence, 2005) and then the fraction of rain over total precipitation during precipitation days (R_{mix}) according to the following equations (Marks et al., 2013; Zhang et al., 2017):

$$R_{mix} = \begin{cases} 0, & \text{if } DPT < -1^{\circ}\text{C} \\ 0.5 + 0.5DPT, & \text{if } -1^{\circ}\text{C} \leq DPT \leq 1^{\circ}\text{C} \\ 1, & \text{if } DPT > 1^{\circ}\text{C} \end{cases} \quad (1)$$

This approach to phase partitioning assumes dew-point temperature to be closely related to the temperature at which precipitation forms, hence to its phase. Note that this method only allows one to estimate the proportion of rain over total precipitation, and not the rain and snow totals. Observed R_{mix} during precipitation days was compared to the corresponding PRMS variable hru_rain/hru_ppt. Both R_{mix} and hru_rain/hru_ppt are thus nondimensional and between 0 (no rain, all snow) and 1 (no snow, all rain). While air temperature could also be employed to estimate precipitation phase, Marks et al. (2013); Jennings et al. (2018) showed that dew-point temperature performs better as an indicator of phase.

Because of the large extent and elevation range of each HRU in the model, comparing measurements located within a given HRU with simulations at the corresponding centroid could lead to spurious results. In order to compare measurements with topographically representative locations, we collected all centroids located within $\pm$ 0.1° in latitude, $\pm$ 0.1° in longitude, and $\pm$ 152 m in elevation from each source of measurements.

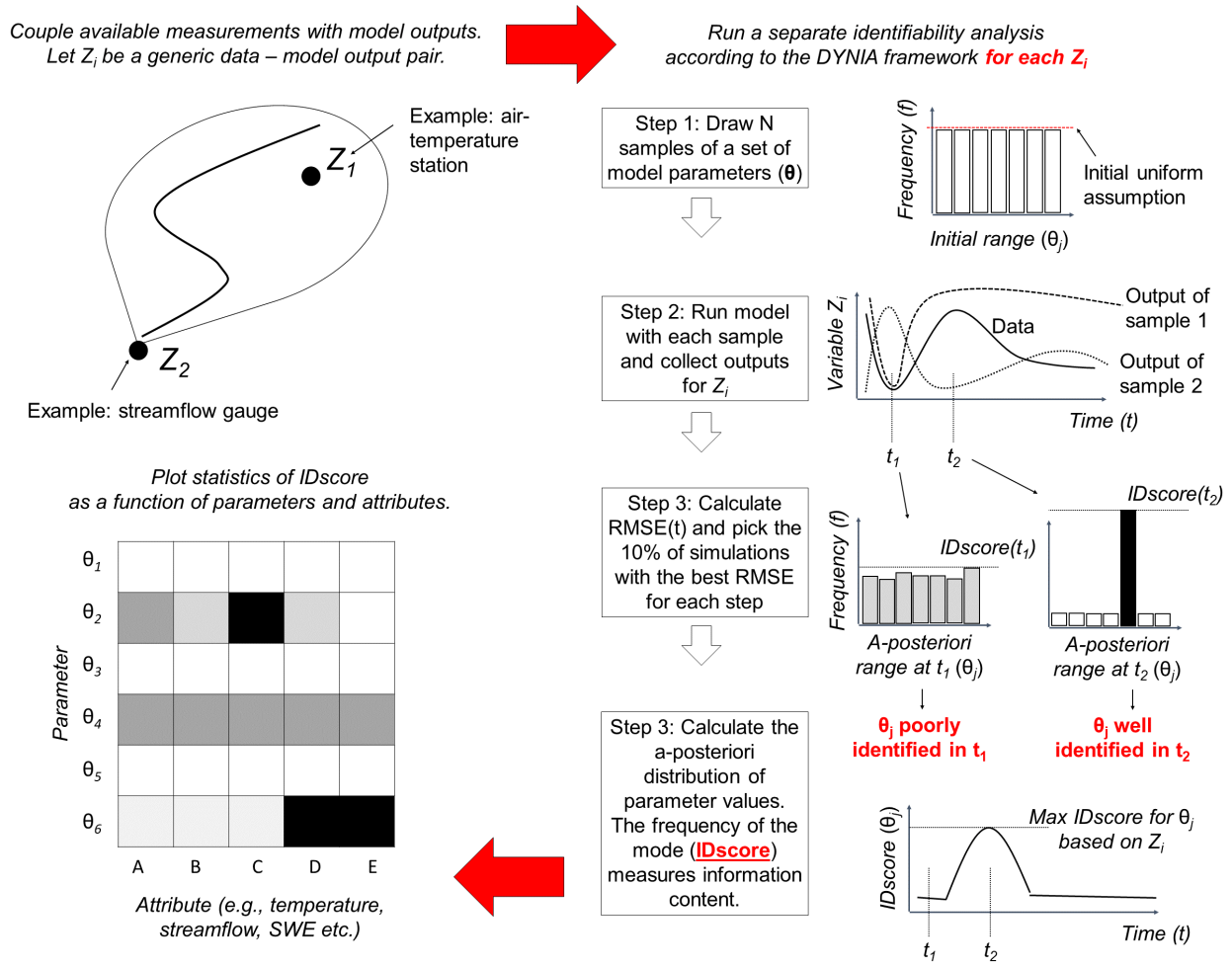

Figure 2: Dynamic Identifiability Analysis (DYNIA) using multiple attributes (see Section 2 for details).

We then selected the centroid with the smallest difference in aspect from the measurement location as its modeling counterpart because elevation and aspect are the main drivers of hydrologic processes in mountain areas (Bales et al., 2006; Oroza et al., 2016). Two important exceptions were streamflow (measurements simply compared with corresponding simulations at the outlet of the basin) and SWE, for which we aggregated observed and simulated time series from all centroids in ten elevation bands according to equally spaced percentiles of centroid elevations. Observed time series of SWE for each centroid were derived from the reanalysis product (see Section 2.1) using bilinear interpolation. Approximate coordinates for all weather stations were obtained from CDEC; aspect was estimated using a ∼30-m DTM from the National Elevation Dataset (`https://www.usgs.gov/core-science-systems/national-geospatial-program/national-map`, visited July 15 2019).

2.3.2. Step 2 and 3

Step 2 consisted in a separate identifiability analysis for each of these Z_i and for a given group of parameters (θ), which implicitly defined a prior parameter range. The basic workflow was the same as DYNIA (Wagener et al., 2003), but it was recursively applied to all Z_i. The first task was to choose a group of model parameters to identify and stochastically draw N = 20,000 samples of these parameters from the prior parameter range using the Latin Hypercube sampler in SAFE with five iterations (Pianosi et al., 2015). Table 2 reports the

list of the 33 parameters involved in this study, the corresponding module in the model, and their dimension. We selected these parameters by both building on a preexisting list by Markstrom et al. (2016) and leveraging expert knowledge of PG&E forecasters on the river.

These parameters cover the entire range of processes in PRMS. Because the focus of this paper is on identifiability, we did not consider parameters that could be calibrated based on topography, canopy properties, or a-priori information (for example, parameters related to snow density or settling, see De Michele et al., 2013), as well as parameters that would be difficult to identify as usually hidden by more dominant processes (for example, the sink term in groundwater flow). We also did not consider any parameter related to initial conditions, because the three initial years of all simulations were discarded from this analysis as a spin-up period.

For parameters with dimensions equal to the number of HRUs in the model ("*nhru*", "*nssr*", or "*ngw*" - see Table 2 for definitions), we assumed the randomly drawn parameter value to be the mean of a normal distribution. We then synthetically reconstructed spatial variability in parameter values across the East Branch by randomly extracting one parameter value for each HRU from this normal distribution (coefficient of variation equal to 0.1). While this method attempted to take into account the spatial variability embedded in parameters with dimension *nhru*, *nssr*, and *ngw*, it only allowed us to draw significant conclusions concerning the basin-wide mean. Multiple options could serve the purpose here and a normal distribution was chosen because it is numerically easy to handle. Note that many parameters with dimension "*nhru*", such as those related to canopy interception, could be estimated a priori, even if they are in practice often tuned on streamflow.

After drawing all samples, PRMS was run using each parametric sample (or realization) at a time; residuals between measurements and the corresponding model outputs were then calculated for all parameter sets and for each Z_i using Root Mean Square Error (RMSE). Rather than calculating a global RMSE over the whole simulation horizon, DYNIA calculates residuals between measurements and paired simulations over a moving window around each time step (width 2W + 1). We used a moving window of 41 days (W = 20), which is comparable to the time scale of monthly forecasts performed by DWR during winter and spring (Rosenberg et al., 2011). The only exception was precipitation-phase partitioning based on dew-point temperature, for which we assumed W = 1 to take into account intermittency of precipitation (De Michele & Ignaccolo, 2013). For the same reason, precipitation phase was the only variable for which local RMSEs were calculated even if some days within the evaluation window had no evaluation measurement (that is, no precipitation). The evaluation period went from October 1, 2004 to September 30, 2017 and thus covered the most recent multi-year California drought (Bales et al., 2018). Three years of spin up (October 2004 to October 2007) were removed from all results.

These temporal trajectories of local residuals were leveraged to detect periods of high information content for a specific parameter, henceforth *dynamic information content*. To this end, we selected a subset of simulations with the best RMSEs for each time step and each Z_i (*behavioral solutions*, see Beven, 2001). This selection was performed by ranking parameter samples for each time step based on RMSEs and choosing the subset composed by the best 10%. Each of these *behavioral simulations* was obtained using a specific value of each parameter, meaning that an *a-posteriori* distribution of parameter values for behavioral solutions could be built for each time step, each parameter, and each Z_i. In doing so, all solutions were assumed to be equally likely and no likelihood weighting was performed (following the implementation of DYNIA in SAFE, Pianosi et al., 2015). The *a-posteriori* probability distributions of parameter values was approximated using histograms with 15 bins.

Simulation periods during which the *a-posteriori* distribution is unimodal and this mode has a relatively high frequency are information-rich periods for a given parameter and a specific measurement-state variable pair Z_i (see schematic in Figure 2). This means that measuring that environmental attribute at location i and during those periods is particularly informative about the optimal value of that parameter with regard to simulating the state variable stored in Z_i. We measured dynamic information content using the frequency of the mode of the *a-posteriori* probability distribution (henceforth, defined as IDscore). The IDscore is thus adimensional and always between 0 and 1. Because one *a-posteriori* distribution was available at each time step, temporal trajectories of IDscore were computed to better identify local maxima and thus

Table 2: List of the thirty-three PRMS parameters considered in this work, their corresponding module in the model, and their dimension (see Section 2.2 and Section 2.3 for an explanation of these modules). Dimension "*one*" means that the parameter has neither spatial nor temporal variability; dimension "*nmonths*" means that the parameter has temporal (monthly) but no spatial variability; dimension "*nhru*" means that the parameter has spatial, but no temporal variability (one value per HRU). Subsurface and groundwater parameters have dimensions "*nssr*" and "*ngw*" (number of subsurface and groundwater reservoirs, respectively). In the implementation of PRMS on the East Branch, $nssr = ngw = nhru$ (Koczot et al., 2004).

Parameter	Module	Dimension
tmax_allsnow	Precipitation Distribution	*one*
tmax_allrain	Precipitation Distribution	*nmonths*
adjmix_rain	Precipitation Distribution	*nmonths*
tmax_lapse	Temperature Distribution	*nmonths*
tmin_lapse	Temperature Distribution	*nmonths*
dday_slope	Radiation distribution	*nmonths*
dday_intcp	Radiation distribution	*nmonths*
radj_sppt	Radiation distribution	*one*
radj_wppt	Radiation distribution	*one*
radadj_slope	Radiation distribution	*one*
radadj_intcp	Radiation distribution	*one*
ppt_rad_adj	Radiation distribution	*nmonths*
tmax_index	Radiation distribution	*nmonths*
radmax	Radiation distribution	*one*
jh_coef	Potential ET	*nomnths*
potet_sublim	Potential ET	*one*
snow_intcp	Canopy Interception	*nhru*
srain_intcp	Canopy Interception	*nhru*
wrain_intcp	Canopy Interception	*nhru*
cecn_coef	Snow module	*nmonths*
freeh2o_cap	Snow module	*one*
smidx_coef	Surface runoff	*nhru*
smidx_exp	Surface runoff	*nhru*
carea_max	Surface runoff	*nhru*
soil_moist_max	Soil Zone	*nhru*
soil_rechr_max	Soil Zone	*nhru*
soil2gw_max	Soil Zone	*nhru*
ssrcoef_lin	Soil Zone	*nssr*
ssrcoef_sq	Soil Zone	*nssr*
ssr2gw_rate	Soil Zone	*nssr*
ssrmax_coef	Soil Zone	*nssr*
ssr2gw_exp	Soil Zone	*nssr*
gwflow_coef	Groundwater	*ngw*

information-rich periods with regard to a specific parameter, a specific attribute, and a specific Z_i (see again Figure 2). These temporal trajectories were summarized with standard descriptive statistics (temporal mean, maximum, or percentiles of IDscore) to compare information content for different attributes and across different parameters and locations (step 3, Figure 2).

Other metrics have already been proposed to quantify dynamic information content. The most popular one is the ratio between the *a-posteriori* confidence interval of a parameter and its prior range (Wagener et al., 2003; Ghasemizade et al., 2017), which provides the same information as the IDscore for the scope of identifiability.

2.4. Evaluation strategy

Results were assessed from three perspectives. First, we quantified the maximum information content of each attribute for identifying a given param-

eter by calculating maximum IDscore across the whole simulation horizon for each attribute, each parameter, and each Z_i separately (see Section 3.1). We then assessed the variability of this maximum IDscore with attribute and with elevation, the main driver of hydrologic processes in mountain watersheds (Bales et al., 2006). We also computed daily medians and daily quartiles of IDscore across all simulation years to assess the temporal variability of information content of each attribute (see Section 3.2). These temporal statistics were computed by merging all results for the same attribute from all measurement locations to create one unique posterior distribution of IDscore values. This computation was restricted to parameters in the model with maximum IDscore larger than 0.4 (henceforth, "well-identified" parameters, see Section 3.1).

Second, we looked at the temporal variability of well-identified optimal parameter values, that is, the values corresponding to the peak in the daily *a-posteriori* probability distribution obtained by DYNIA for each attribute, each parameter, and each Z_i. In the framework of DYNIA, these values (and/or their confidence intervals, see Section 2.3) are the closest parallel to the optimal parameter values that one would estimate through a standard calibration procedure. Temporal variability was assessed by computing DYNIA-based optimal parameter values on a monthly basis (water years 2007 to 2017) using both streamflow and the most informative, alternative attribute (when different from streamflow). In order to estimate optimal parameter values using the most informative days during each month, we first computed the third quartile of daily IDscore for each month and each attribute and collected all daily parameter values for the same month with IDscore greater than or equal to this quartile. We finally computed the monthly 25^{th}, 50^{th}, and 75^{th} percentiles of these subsets across all years (see Section 3.3). We chose a monthly time scale as this corresponds to the operational time scale of water-supply and runoff forecasts in California (Harrison & Bales, 2016).

Third, we assessed the impact on the simulation of the water budget of tuning parameter values based on ground-based attributes. To this end, we first estimated new optimal values for all well-identified parameters as the median of their daily values across the entire identification period (water years 2008 to 2017) that were above a threshold in IDscore of 0.2 (according to the most informative attribute). We chose the median to make our parameter estimates more robust to extremes and more representative of central tendencies. A threshold of 0.2 corresponds to three times the probability of a uniform distribution as approximated by 15 bins and was chosen to detect periods in which information content was reasonably above noise level. We defined a fixed threshold here to use the same threshold across all parameters and attributes.

We then compared simulations using the original parametric set (Koczot et al., 2004) and that estimated through this identification procedure. Our evaluation variables were maximum and minimum air temperature, precipitation-phase proportion as estimated from dew-point temperature (rainfall over total precipitation), solar radiation, SWE, and streamflow, which are the attributes for which we had multi-year data. Changes in performance were quantified using differences in Root Mean Square Error (RMSE) and in modified Kling-Gupta Efficiency (KGE, Kling et al., 2012) between the simulation informed by DYNIA results and the one using the original parameter values (henceforth, Δ). The modified KGE differs from the original KGE in that it uses the ratio of the coefficient of variations of simulations and observations instead of the ratio of their standard deviations to remove the effect of bias (Santos et al., 2018).

RMSE is a direct measure of simulation residuals and is dimensional; as such, it quantifies performance gains or losses in terms that have clear practical implications. On the other hand, KGE better weights multiple components of predictive errors (shape, timing, bias, and variability, see Santos et al., 2018), with the downside that its values have less direct interpretability, similarly to the Nash-Sutcliffe Efficiency (Schaefli & Gupta, 2007). An improved simulation of a given target variable would return a negative ΔRMSE and a positive ΔKGE. To facilitate further discussions, we will consider Δ(1 - KGE), for which negative values indicate an improved simulation using results from DYNIA.

For air temperature, precipitation-phase proportion, and solar radiation, we calculated performance differences for each measurement location across water years 2008 to 2017. For these variables, we used both ΔRMSE and Δ(1 - KGE), even if the latter may not be suitable for strongly bimodal variables like precipitation-phase proportion (see Pool et al., 2018 and Section 4). While this period was also used to identify parameter values, and so this analysis is not a fully independent val-

idation, this is also the only decade for which we had data; assessing model performance over a relatively long period is key given the variability of California's climate. Using any single year of DYNIA results to estimate optimal parameter values showed comparable results in terms of performance differences (not reported for brevity).

For SWE, we computed ΔRMSE for each HRU across water years 1985 to 2006, that is, the maximum timespan of SWE data outside the identification period. We did not compute KGE for SWE as this performance metric is not defined for constant time series (standard deviation equal to zero), and this is the case for low-elevation, snow-free HRUs. To separate the effect of Dynamic Identifiability on RMSE from that of other sources of uncertainty and in particular precipitation, we also computed median April-1 cumulative precipitation for each HRU according to PRMS and compared that to observed precipitation at a collection of ground-based precipitation gauges (stations ANT, CSH, DOY, JDP, KTL, QCY, TAY, and TVL, see Table 1). April 1 is the reference date of peak accumulation in the western U.S. and cumulative precipitation on this date is thus strongly related to SWE (Knowles et al., 2006). This precipitation comparison was performed across water years 2007 and 2017, which is the decade with the largest amount of available precipitation gauges on the river.

For streamflow, we computed annual (water-year) ΔRMSE and Δ(1 - KGE) over the period 1972-2007, the maximum timespan of data outside the identification and a three-year spin-up periods. Contrary to all other evaluation variables, we considered an annual time scale for streamflow to identify differences in performance for different year types as classified by PG&E's license conditions (critically dry, dry, normal, wet, see https://cdec.water.ca.gov/snow/bulletin120/ for some background on year types in California – visited July 14 2019).

3. Results

3.1. Maximum information content of environmental attributes

Seven PRMS parameters out of thirty-three show a maximum IDscore larger than 0.4 for at least one attribute (see Figure 3). This subset represents the seven well-identified parameters by DYNIA: *tmax_lapse* and *tmin_lapse*, *tmax_allsnow* and *tmax_allrain*, *radmax*, *smidx_coef*, and $gwflow\_coef$. The parameters with the highest IDscore (hence, the most identifiable ones) are the two lapse rates for minimum and maximum air temperature, *tmax_lapse* and *tmin_lapse*, both well identified using air temperature measurements at any elevation (maximum IDscore ∼0.67, refer to Figure 3 for all results in this Section). Air temperature provides up to 2.9-3.6 times the information of streamflow for identifying lapse rates.

Air temperature, coupled with relative humidity to estimate dew-point temperature, is also the most informative attribute to identify the maximum-temperature threshold used by the model to classify all precipitation as snow, *tmax_allsnow*. Maximum IDscore for this parameter-attribute combination, however, significantly decreases with increasing elevation (max IDscore ∼0.58-0.63 below 1200 m and 0.24 above 2200 m, see Section 4 for an interpretation). At medium-high elevation (above ∼1700-2000 m), the information content of dew-point temperature for *tmax_allsnow* is comparable to that of SWE (max IDscore of 0.21 ± 0.017 across all elevation bands).

The maximum-temperature threshold to classify all precipitation as rain (*tmax_allrain*) is less identifiable than the one for snow when using dew-point temperature (max IDscore ∼0.54-0.48 below 1200 m and 0.22 above 2200 m, mixed precipitation occurring between this threshold and *tmax_allsnow*). SWE data are slightly more informative for this parameter than they are for *tmax_allsnow* (max IDscore of 0.21 ± 0.022 across all elevation bands). Overall, dew-point temperature provides up to 2.1-2.6 times the information of streamflow for identifying both precipitation phase-partitioning parameters (that is, to detect rain vs. snow).

Parameters related to radiation distribution, canopy interception, evapotranspiration, and snow accumulation and melt are poorly identified by any attribute. Thus, these processes appear to be the most relevant sources of conceptual uncertainty in this model. Among these poorly identified parameters, relative peaks in information content are obtained by radiation to identify *radmax* at all elevations (*radmax* accounts for atmospheric transmissivity in the computation of maximum potential solar radiation), and by soil-water storage to identify jh_coef (a coefficient used in the computation of evapotranspiration). The IDscore for these parameter-attribute combinations is equal to 0.42-0.33 and 0.20, respectively. While streamflow has

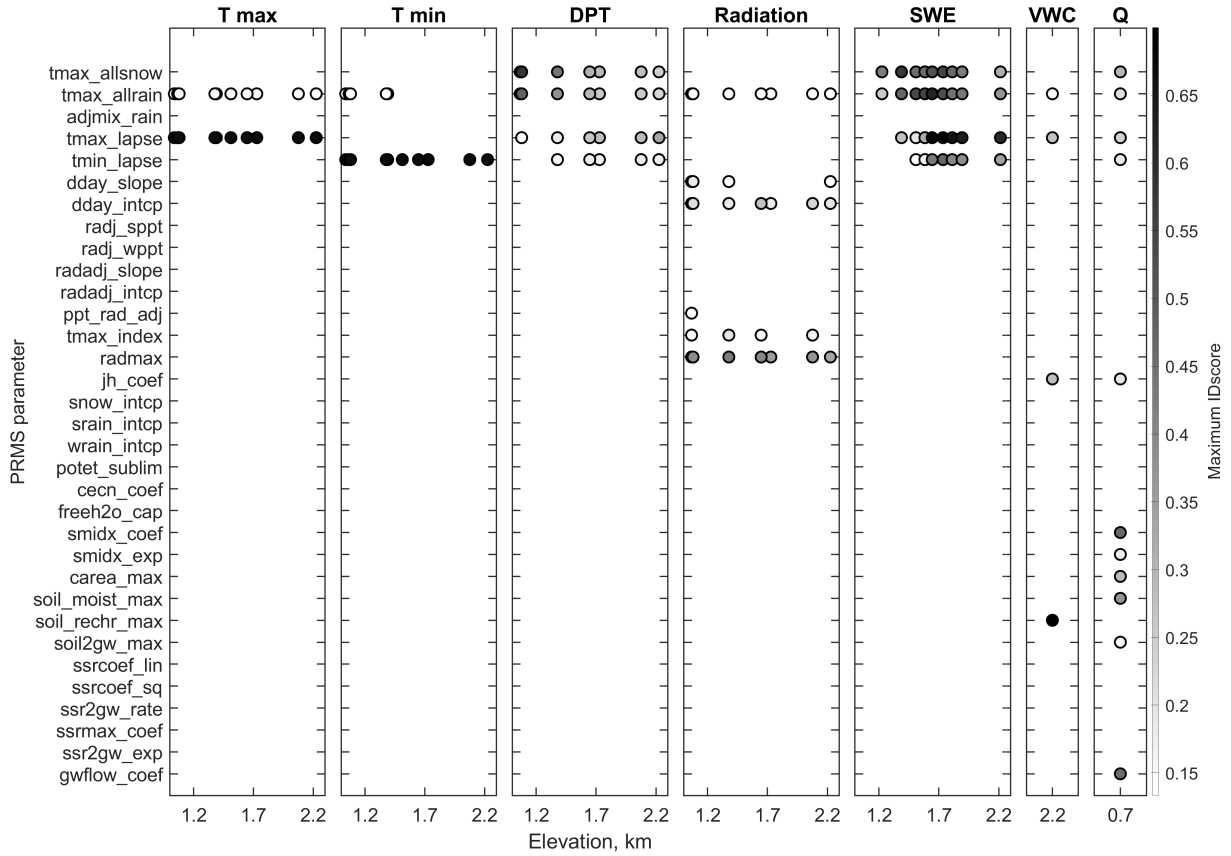

Figure 3: Maximum IDscore for a selection of PRMS parameters and all attributes, plotted as a function of the elevation of the measurement location. IDscore values below ∼0.13 are not reported for readability. Results for snow depth are not reported as IDscore for all parameters and for this attribute are below ∼0.13. The color scale represents IDscore (see colorbar on the right). T max, T min, DPT, Radiation, SWE, VWC, and Q are maximum and minimum air temperature, dew-point temperature, incoming shortwave solar radiation, snow water equivalent, volumetric soil-water storage, and streamflow, respectively.

no information content for *radmax* and, in general, any radiation-related parameter, it returns a comparable IDscore to soil-water storage to identify jh_coef (0.19).

Runoff, surface-soil-moisture, and groundwater parameters are generally more identifiable using streamflow than using any other attribute. Peaks in information are obtained for $smidx\_coef$ and $gwflow\_coef$ (maximum IDscore ∼0.46 for both), two parameters related to surface-runoff generation and baseflow, respectively (see Markstrom et al., 2015 for details). Measurements of soil-water storage provide some information for recharge-moisture saturation (maximum IDscore of 0.31). No parameter related to the subsurface reservoir can be identified from any attribute.

3.2. Seasonal patterns of information content

Both lapse rates show a consistently high IDscore throughout the year when identified using temperature (see Figure 4 for all results in this Section). No peak or seasonality in information content is evident; variability across years is also relatively small (the first and third quartiles of IDscore are systematically between 0.5 and 0.75). Despite being the most identified parameters by DYNIA, IDscore for lapse rates is still smaller than 1 (see Section 4.1 for a discussion of this outcome).

The IDscore for phase-partitioning parameters ($tmax\_allsnow$ and $tmax\_allrain$) is generally higher when using dew-point temperature than when using streamflow, as well as in winter than in summer (when trajectories according to the two attributes overlap). Again when using dew-point temperature, both day-to-day and inter-annual

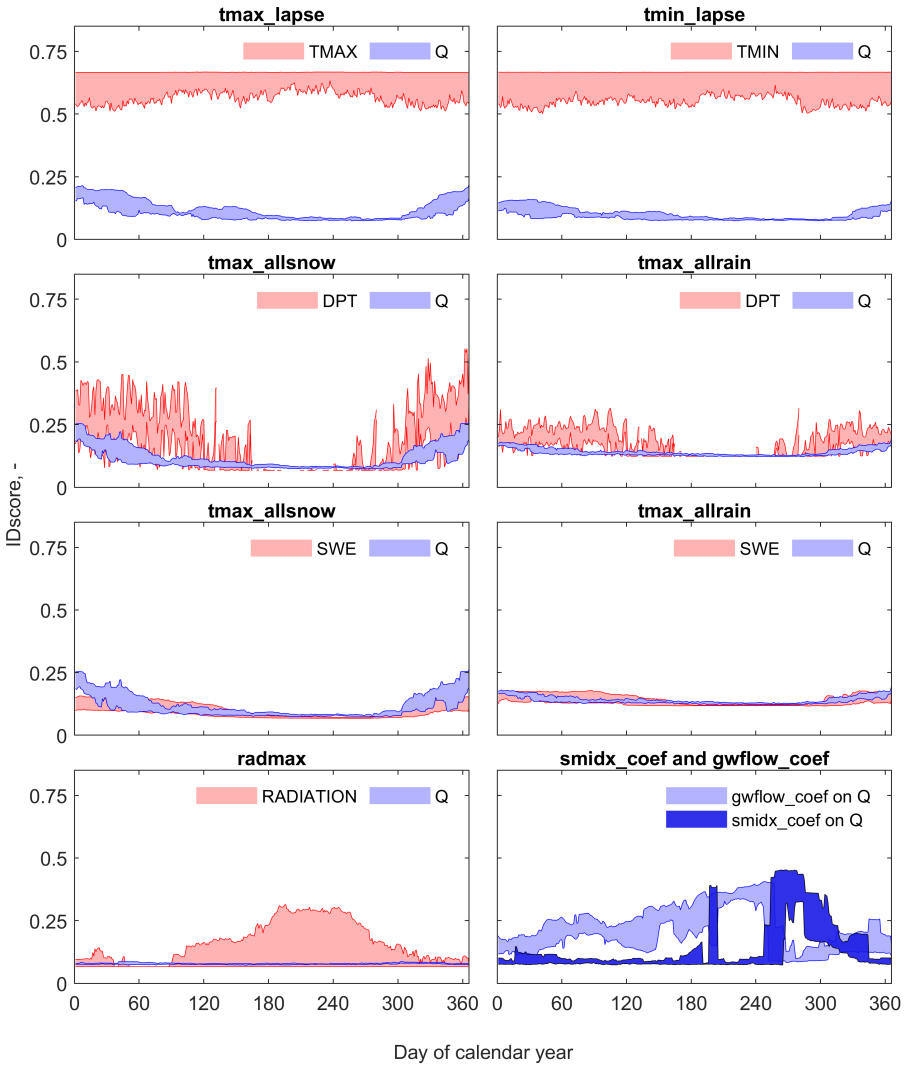

Figure 4: Daily first and third quartile of IDscore across all simulation years for the seven well-identified parameters in PRMS (quartiles are reported as a colored range). For parameters that were better identified on alternative attributes than on streamflow (see Figure 3), we compare statistics of IDscore between these alternative attributes and streamflow. TMAX, TMIN, Q, DPT, SWE, and RADIATION refer to (daily) identification attributes of air maximum temperature, air minimum temperature, streamflow, dew-point temperature, snow water equivalent, and incoming shortwave solar radiation, respectively.

variability in IDscore are large due to intermittency of precipitation events. For instance, the third-quartile IDscore across consecutive days can range from ∼0.40 to ∼0.30 depending on when precipitation occurred (low-precipitation days being less informative, hence a low IDscore). The IDscore for *tmax_allsnow* using dew-point temperature is a little higher than the IDscore for *tmax_allrain* (∼0.3 vs. ∼0.2). The IDscore for both parameters identified using SWE across all elevation bands is comparable to that obtained using streamflow, with just a small gain of information for *tmax_allrain*. All these outcomes agree with results in Section 3.1.

All other well-identified parameters show clear seasonal patterns in information content: IDscores for *radmax* (radiation distribution) and

smidx_coef (surface runoff) peak in summer and fall, respectively; the IDscore for *gwflow_coef* (groundwater flow) increases from January to August-September and then quickly decreases. This drop is consistent across years and occurs simultaneously to a rise in information content for *smidx_coef* (surface runoff), which is compatible to a transition in dominant processes from summer baseflow to rainfall-runoff during fall. This finding is evidence that DYNIA correctly identifies dominant processes and how they affect information content across different attributes and across the water balance. The quartiles of *radmax* tend to be relatively large during summer because information content for this parameter strongly depends on daily cloud-cover conditions.

3.3. Variability of optimal parameter values

Monthly optimal values for precipitation phase-partitioning parameters (*tmax_allsnow* and *tmax_allrain*) display much smaller inter-annual variability when identified on dew-point temperature than on streamflow, especially during periods of relatively high information content for the former (January to April and then October to December, see Figure 5). While optimal values for *tmax_allsnow* do not show any interannual variability for any elevation during periods of relatively high IDscore, optimal values for *tmax_allrain* are more variable - especially at lower elevations (∼1000 m). Both parameters show significantly different values during periods with low information content (summer) compared to high-information periods like winter. While the obtained optimum values for *tmax_allrain* may in principle seem very high (∼ 15°C), PRMS classifies all precipitation as rain only if *maximum daily* temperature is greater than *tmax_allrain*. This threshold is thus not directly related to air temperature during a precipitation event (see Section 4.1 for a more extensive discussion).

Optimal values for lapse rates (identified on air temperature) and *radmax* (identified on solar radiation) show marked intra-annual variability across all elevations (see Figures S1 and S2). It follows that, for these parameters, a relatively high monthly IDscore does not guarantee the same small temporal variability observed for precipitation phase-partitioning parameters using dew-point temperature. The range of inter-annual variability using alternative attributes (air temperature and solar radiation) is, however, smaller than that obtained using streamflow as the identification source.

Parameter *smidx_coef* shows large temporal variability when considering periods with relatively low information content and no variability during period of high IDscore, while parameter *gwflow_coef* shows no variability in optimal parameter values across the whole year regardless of IDscore (both parameters are well identified by streamflow, see Figures S3 in the Supporting Information). Overall, results in this Section thus showed that temporal variability in optimal parameter values is only weakly coupled with IDscore, even if relying on the most informative attribute generally minimizes this variability.

3.4. Impact on water-budget predictability

Tuning well-identified parameter values based on DYNIA improves simulations of air temperature (see Figure 6), especially at medium-high elevations. Improvements are larger for minimum than for maximum temperature (ΔRMSE up to ∼ -3 °C for minimum temperature at KTL, the highest station in the basin; however, only one year of data is available). Results are consistent when comparing ΔRMSE and Δ(1 - KGE), but the latter is more frequently positive than the first (2 vs. 0 cases for maximum air temperature, 3 vs. 1 cases for minimum temperature, a positive difference meaning a worse performance after Identifiability Analysis), likely because RMSE was used as the objective function in DYNIA.

The well-identified set of parameters also improves simulations of precipitation-phase partitioning in terms of ΔRMSE, especially at low elevations where precipitation-phase parameters are more identifiable (see Figure 3). Performance changes in terms of Δ(1 - KGE) are generally positive (that is, a worse performance after Identifiability Analysis), but phase-partitioning data are strongly bimodal and this challenges the use of KGE (Pool et al., 2018). In terms of daily average solar radiation, ΔRMSE is low compared with peaks in annual solar radiation (up to -10 W/m^2 compared to peaks of 350-500 W/m^2 in summer or 100-150 W/m^2 in winter); changes in terms of Δ(1 - KGE) for this variable are mostly negative (improved performances), but also relatively small.

Results for SWE depend on elevation (correlation coefficient equal to -0.4036, Figure 7), with performance predominantly improved at elevations between ∼1600 and 1800-2000 m (ΔRMSE up to -42

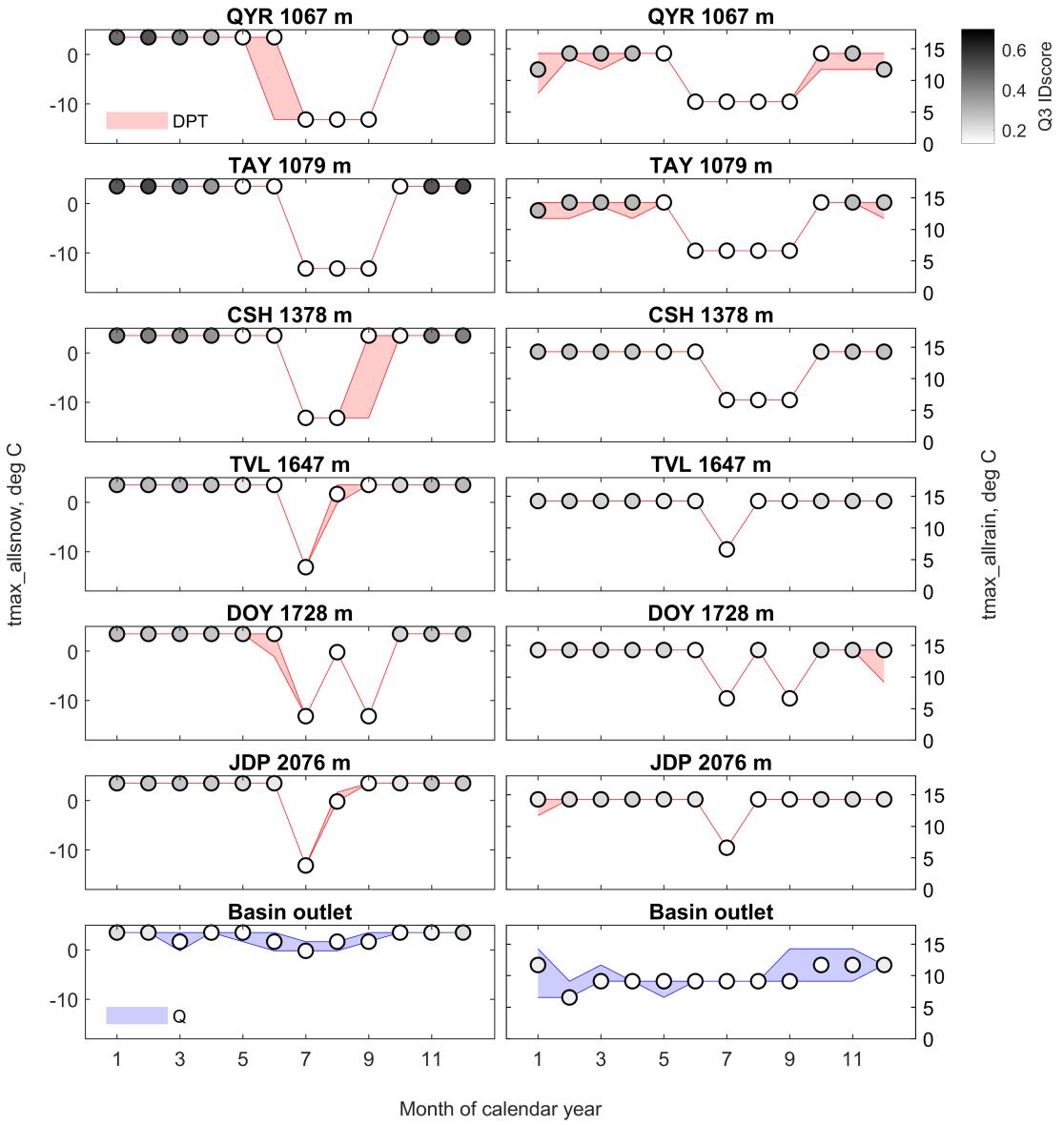

Figure 5: Monthly optimal values for the two parameters that PRMS uses for precipitation-phase partitioning (*tmax_allsnow* and *tmax_allrain*) as identified by dew-point temperature measurements at different elevations and by streamflow at the basin outlet. The colored range represents the (monthly) first and third quartiles of optimal parameter values across all years in the identification period (water years 2008 to 2017); the circles represent monthly medians. The color scale of each circle represents the third quartile of monthly IDscore (see colorbar in the upper right corner). DPT and Q are dew-point temperature and streamflow, respectively.

mm), where the range of spatial variability in simulated precipitation according to PRMS includes the values independently measured by ground-based gauges. At elevations below 1600 m, ΔRMSE increases up to +80 mm; at these elevations, however, PRMS significantly overestimates April-1 precipitation compared to gauges. Above 2000 m, results are inconclusive because both the number of HRU centroids and that of precipitation gauges decrease.

The well-identified parametric set generally af-

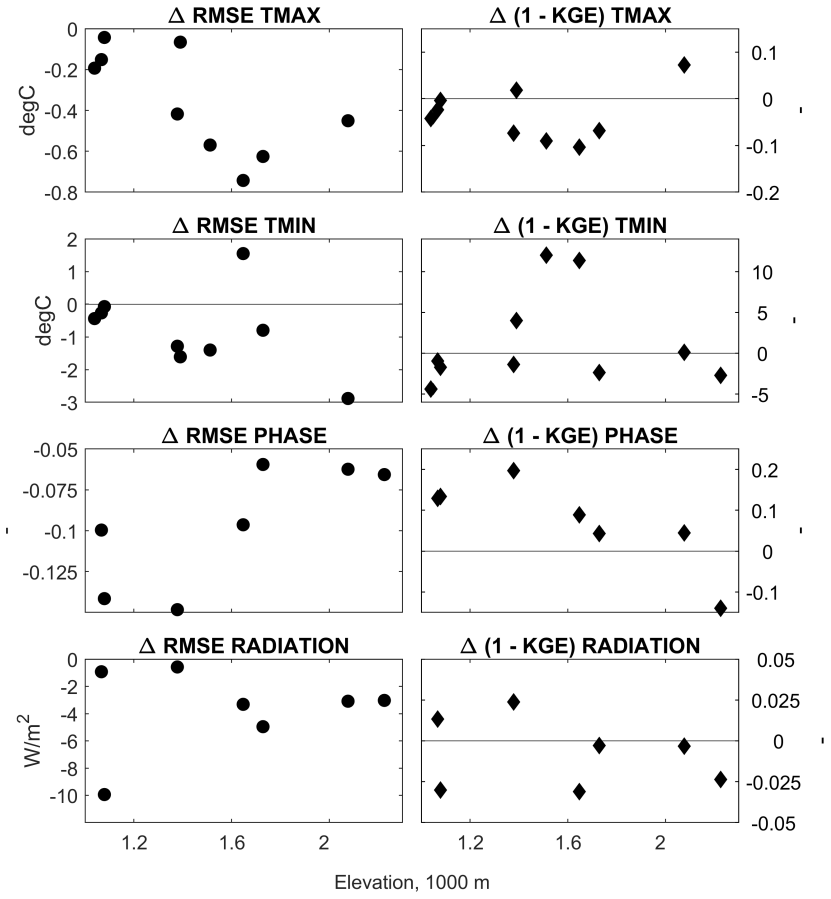

Figure 6: Difference in Root Mean Square Error (ΔRMSE, left) and Kling-Gupta Efficiency (Δ(1 - KGE), right) between simulations of maximum and minimum daily air temperature (TMAX and TIM), precipitation-phase partitioning (PHASE), and incoming shortwave solar radiation (RADIATION) obtained by tuning optimal parameter values based on DYNIA and those obtained by using the original parametric set (Koczot et al., 2004). The evaluation period for all these attributes is the same as the identification period (2008-2017) due to data availability. Each point represents one ground-based measurement location (see Table 1), plotted as a function of its elevation. The horizontal line denotes $\Delta = 0$. With regard to PHASE, the evaluation is performed comparing the proportion of rain over total precipitation as estimated by PRMS with observations obtained by partitioning precipitation phase during precipitation days using dew-point temperature.

fects both accumulation and melt seasonal patterns, but no clear change in peak-SWE timing is observed (Figure S4). The overall impact of tuning lapse rates (*tmax_lapse* and *tmin_lapse*), phase-partitioning parameters (*tmax_allsnow* and *tmax_allrain*), and *radmax* (the only well-identified parameters that have a direct impact on SWE) is a topographic shift in snow accumulation from higher to lower elevations. The average difference in simulated basin-wide SWE between the well-identified and the original parametric sets is equal to -4 mm, that is, 4.3% of average, simulated basin-wide SWE using the original parameters.

Results for streamflow are inconsistent, with ΔRMSE spanning -5 m^3/s and +15 m^3/s and Δ(1 - KGE) spanning -0.07 and +0.75 (Figure 8). In 30 out of 46 years (65%), ΔRMSE is larger than 0, meaning that tuning only well-identified parameters generally decreases the performance of PRMS for streamflow (31 out of 46 for Δ(1 - KGE)). No evident correlation emerges between the sign of ΔRMSE and Δ(1 - KGE) and year types: for example, negative ΔRMSE include both wet (1993) and critically dry years (1989).

From a seasonal standpoint, changing well-identified parameters in PRMS affects both spring

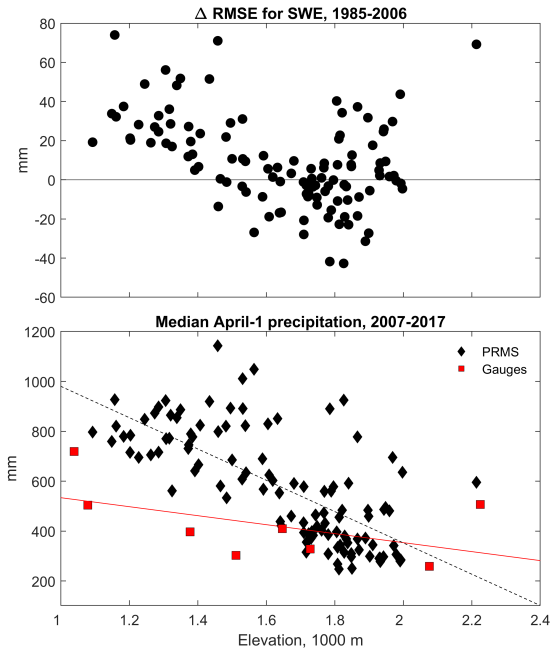

Figure 7: Top panel: difference in Root Mean Square Error (ΔRMSE) between simulations of snow water equivalent (SWE) obtained by tuning optimal parameter values based on DYNIA and those obtained by using the original parametric set (Koczot et al., 2004) over water years 1985-2006. Each point represents ΔRMSE for one HRU centroid, plotted as a function of its elevation. The horizontal line denotes $\Delta = 0$. Bottom panel: median April-1 precipitation simulated by PRMS for each HRU centroid over the period 2007-2017 (water years) compared to that observed by precipitation gauges over the same period (both plotted as a function of station/centroid elevation). Lines represent linear regressions between April-1 precipitation and elevation according to the two datasets (dashed is for PRMS, solid is for precipitation gauges).

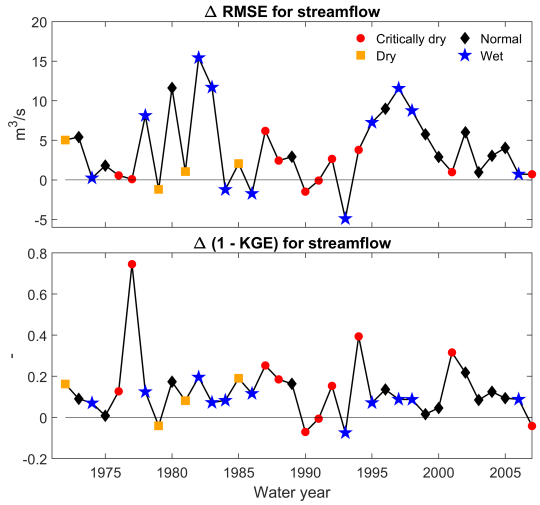

Figure 8: Annual (water-year) differences in Root Mean Square Error (ΔRMSE) and Kling-Gupta Efficiency (Δ(1 - KGE)) between simulations of streamflow (Q) obtained by tuning optimal parameter values based on DYNIA and those obtained by using the original parametric set (Koczot et al., 2004). Colors and symbols refer to water-year types (see Section 2.4). The evaluation period includes water years 1972 to 2007. The horizontal line denotes $\Delta = 0$.

and summer predictions (Figure 9) as well as the overall water balance. The most evident effects are a slight decrease in streamflow in early spring (around water-year day 100), an increase in late-spring flow (between water-year days 100 and 200), and a small decrease in summer flow (after water-year day 250). These effects are related to larger snow melt occurring at low elevations with the well-identified parameter set (Figure S4). This shift from rain to snow at low elevations increases overall runoff compared to the original simulation (relative difference between cumulative mean daily streamflow before and after DYNIA equal to $\sim +8.7\%$ compared to the original parametric set).

4. Discussion

4.1. Implications

The most important finding of this paper is that spatially distributed, ground-based measurements provide added information for identifying hydrologic-model parameters, since they allowed us to identify more parameters than streamflow alone (see Figure 3). This supports previous evidence that collecting more measurements directly informs parameter calibration in hydrologic models (see for example Vrugt et al., 2002; Fenicia et al., 2008; Wanders et al., 2014; Silvestro et al., 2015; Saksa et al., 2017; Nijzink et al., 2018). Even though the formulation of DYNIA is independent of attribute (Wagener et al., 2003), streamflow has been the most frequent choice to date for Identifiability Analysis of hydrologic models (Wagener et al., 2003; Pianosi & Wagener, 2016; Abebe et al., 2010; Hoey et al., 2015). More recently, Ghasemizade et al. (2017) combined DYNIA and sensitivity analysis using measurements of soil-water storage, but focused on a very local scale rather than a full hydrologic model. The present study complements these efforts by (1) investigating the dynamic information content of spatially distributed, mostly ground-based attributes and (2) focusing on mixed

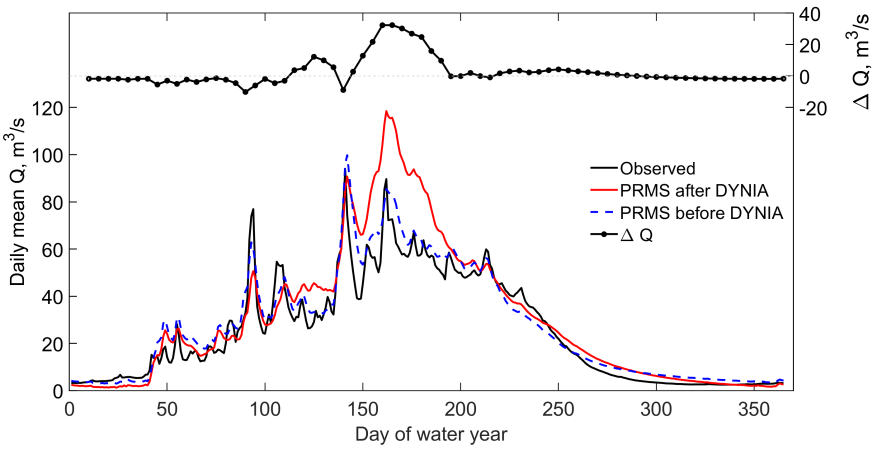

Figure 9: Observed versus simulated long-term mean streamflow at the outlet of the East Branch of the Feather River (water years 1972 to 2006 – left y axis). Simulated results were obtained by both running PRMS as is and tuning well-identified parameters using results from DYNIA (see Section 2.4 for details). The black dotted line represents differences between these two simulations (simulation after DYNIA minus simulation before DYNIA, ΔQ – right y axis).

rain-snow mountain catchments, which introduce large predictive uncertainty in hydrologic models (Rössler et al., 2014; Harpold et al., 2017).

One of the most relevant sources of this predictive uncertainty is spatially distributed precipitation-phase partitioning, that is, the distribution of rain vs. snow across the landscape. The development of new source of phase-partitioning data and their application in predictive models have recently been suggested as an urgent research need by Harpold et al. (2017), who pointed to humidity as a potentially valuable, but underutilized source of information for hydrologic models. By showing that coupled measurements of temperature and relative humidity provide peaks in information content to constrain the optimal values of *tmax_allsnow* and *tmax_allrain*, the two parameters used by PRMS to detect rain vs. snow (see Figures 3 and 5), we implicitly validated this hypothesis. The use of dew-point temperature for detecting precipitation phase has been already proposed by Marks et al. (2013), while Zhang et al. (2017) have shown that precipitation phase estimated with the same method we used here tallies with co-located snow measurements. Compared to Marks et al. (2013) and Zhang et al. (2017), the novel contribution of our work is that we employed measurements of dew-point temperature to identify parameters of a hydrologic model. While a few hydrologic models already combine air temperature and relative humidity to classify precipitation as rain or snow (e.g., see Froidurot et al., 2014; Laiolo et al., 2014), most models still rely on ill-defined relationships with air temperature to predict precipitation phase (Harder & Pomeroy, 2014; Harpold et al., 2017).

In addition to being a more informative calibration source, precipitation-phase parameter values identified using dew-point temperature show much less temporal variability than those that were identified on streamflow (Figure 5). This has promising implications for operational forecasters, who lack evaluation data for their predictions and so would be challenged by temporal variability in parameter values (Pagano et al., 2014). Identifiability of phase-partitioning parameters decreases with two separate factors: (1) increasing elevation, and (2) when detecting rain rather than snow (see Section 3.1). The decrease in identifiability with increasing elevation could be because hydrologic models use surface measurements to determine phase, whereas actual phase of precipitation depends on thermodynamic conditions across the atmospheric column and on complex feedback mechanisms with local topography (Harpold et al., 2017). Valley, flat stations (such as QCY, QYR, or TAY in our case) are thus ideal locations to rely on surface data for phase partitioning compared to stations at high elevations and in complex terrain. The decrease in identifiability with increasing proportion of rain, on the other hand, agrees with previous findings by Harder & Pomeroy (2014), who showed that uncertainty in precipitation-phase partitioning increases

with warmer conditions.

Distributed ground-based measurements also provide significantly higher information than streamflow for calibrating lapse rates (*tmax_lapse* and *tmin_lapse*, see Figure 3) and so for constraining uncertainty related to air-temperature distribution, which directly impacts phase partitioning, evapotranspiration, and snow-melt patterns (Markstrom et al., 2015). This result is not surprising, as linear relationships between elevation and air temperature have been already proposed (Rolland, 2003) and have shown the same marked temporal variability that we obtained here (Figure S1). As a result, similar linear relationships between temperature and elevation are a frequent predictive approach in hydrologic models employed in mountain regions (see e.g. Schaefli et al., 2005; Bongio et al., 2016).

It is more surprising that the IDscore of temperature lapse rates never reached the theoretical upper bound of 1, despite these lapse rates being the most identified parameters in this work (Figure 4). We interpret this as due to *tmax_lapse* and *tmin_lapse* not being able to fully reproduce the day-to-day and inter-annual variability in temperature gradients, likely because they are specified as monthly parameters. Indeed, Root Mean Square Errors of the well-identified lapse rates for daily maximum and minimum temperature are smaller than those of the original parametric set (Figure 6), but are still on the order of $\sim$2 to 6°C at seasonal time scale – with peaks for minimum temperature (Figure S5). Variability in temperature seems particularly challenging to capture for remote locations that are separated by topographic barriers from the input station of QCY-QYR (Figure 1), which is used by the model as a seed to distribute temperature to all HRUs (see the cases of JDP and ANT in Figure S5). This difficulty in correctly reproducing temperature spatial variability may also explain why optimal values of *tmax_allsnow* and *tmax_allrain* in Figure 5 are higher than the temperature range at which mixed precipitation tends to occur (between -0.4 and 2.4°C in 95% of the cases in the global assessment by Jennings et al., 2018): DYNIA may have partially optimized these parameters to compensate for biases in reproducing air temperature. Dynamic lapse rates or more sophisticated temperature-distribution approaches might be more suitable in such complex terrains (Frei, 2014).

Several intermediate processes in the water budget like radiation distribution, interception, evapotranspiration, snow accumulation and melt, and, partially, soil-water storage are poorly identified even when using potentially more appropriate identification attributes than streamflow such as incoming shortwave radiation, SWE distribution or soil-water storage. The only identifiable parameter in this context, *radmax*, has a very weak impact on hydrologic predictions (see Figure 6). This is clear evidence that the most relevant sources of structural uncertainty in this model lie between the distribution of surface inputs (lapse rates and precipitation-phase partitioning parameters) and soil-to-runoff streamflow generation (*smidx_coef* and $gwflow\_coef$).

While hydrologic models in complex terrain now tend to ingest solar radiation as an external input rather than simulating it, prediction of interception, evapotranspiration, and snow accumulation and melt is still internally performed (see e.g. Schaefli et al., 2005; Lehning et al., 2006; Silvestro et al., 2018; Markstrom et al., 2015; Bongio et al., 2016) and represents a key driver of water supply in mixed rain-snow regions (Bales et al., 2006, 2018). This finding has thus general implications for distributed models with parameters covering the entire spectrum of hydrologic processes in headwater catchments (Bartolini et al., 2011; Schaefli et al., 2014; Bongio et al., 2016). Most of these parameters were formulated based on local-scale process understanding, but hydrologic models aim at processes acting at much larger scales (Blöschl & Sivapalan, 1995; Pagano et al., 2014). As a result, parameters in hydrologic models like PRMS often lose a clear physical meaning and their optimization mainly depends on fit on data. While this is to our knowledge the first application of DYNIA for such a model (30+ selected parameters out of 100+), even relying on attributes that are explicitly simulated through these poorly identified parameters could only partially overcome the inherent equifinality of these models (Beven & Freer, 2001). A possible explanation is that parametric and conceptual uncertainty are only two out of several predictive-uncertainty sources in hydrologic models, some others being input and output uncertainty (Renard et al., 2010), and separating the role of these uncertainty sources is challenged by the large amount of parameters (and so degrees of freedom) of these models.

A good example from this perspective is the different effect of DYNIA on SWE simulations between low and mid to high elevations (Figure 7),

which is the combination of two sources of uncertainty acting at scales that are beyond the capabilities of DYNIA. On the one hand, no parameter related to the snow module could be identified based on any attribute, which means that the simulation of snow patterns is inherently uncertain in this model as discussed in the previous paragraphs. On the other hand, precipitation in this model is estimated *a priori* using an independent procedure based on average monthly precipitation and *in-situ* measurements at ten stations (see Section 2.2). The accuracy of this procedure is low at low elevations (Figure 7) because most of these ten stations lie on the wet side of the Feather River rather than within the drier East Branch (see Koczot et al., 2004 for extensive details on this). We did not explicitly assess the uncertainty associated with this distribution scheme in this paper because doing so would in principle require recalibrating the model, as discussed by Harder & Pomeroy (2014). While results in this paper have a general validity that goes beyond the specific distribution of precipitation, they also demonstrate that more measurements and improved process representation are both highly needed to effectively constrain predictive uncertainty in distributed-parameter mountain-hydrology models.

4.2. Limitations and future research

A first limitation of this work is that tuning parameter values based on multiple attributes improved predictive performance for the surface water budget, but decreased overall performance for other key predictive variables like SWE (see Section 4.1) and streamflow (see Figure 8). Besides the prominent role played by conceptual and parametric uncertainties in distributed hydrologic models (see Section 4.1), another reason for this drop in performance is that multi-attribute Dynamic Identifiability Analysis is essentially a single-objective procedure, which does not fully account for interactions across parameters and multiple objectives (Zhang et al., 2018). DYNIA has typically been used to diagnose parametric and conceptual uncertainty rather than to directly perform calibrations (Wagener et al., 2003), meaning a full validation of DYNIA-based predictions was beyond the scope of this work. We stress, however, that validation is a necessary step in model development; more extensively assessing the implications of this work for hydrologic predictions in addition to parameter identification is an important target for future work.

That said, applying DYNIA to multiple attributes did provide valuable information to tune specific parameters and so improve the simulation of processes within the integrated water budget that may be misrepresented by a streamflow-based calibration (Figure 6). From this perspective, multi-attribute DYNIA could be leveraged as a first step in a multi-objective calibration. Nijzink et al. (2018) have recently found that soil-moisture and storage-anomaly maps provide valuable information for lumped, conceptual models. Merging information from ground-based and remote-sensing attributes to better constrain parameter identification of distributed-parameter hydrologic models represent another important future direction for this work.

We also used only one objective function to measure information content (RMSE), following the original application of DYNIA in Wagener et al. (2003). A key reason for this decision is that DYNIA computes performances over relatively short time windows compared to those used by standard calibration procedures (in our case, 41 and 3 days for all attributes and dew-point temperature, respectively). The sample size of simulation residuals may thus be too small for statistically more complex performance criteria like KGE, which is also not defined for constant time-series and implicitly assumes normality of data (see Section 2.3 and Pool et al., 2018). By applying DYNIA on streamflow data using KGE, Nash-Sutcliffe Efficiency (NSE), bias, and NSE with logarithm transformation rather then RMSE, we nevertheless found that identification results are quite robust to the objective metric used (Figure S6). The largest deviations compared to RMSE-based IDscore were found for surface-runoff parameters, which was expected given that different performance metrics are sensitive to different characteristics of streamflow patterns (Santos et al., 2018). The IDscore for RMSE and NSE is the same for all parameters, since the latter is a normalization of the first with respect to observed variance.

Finally, we considered a relatively small number of model runs (20,000) compared to standard rules of sensitivity analyses, which, for example, dictate that this number should be up to 1000 times the number of parameters involved (in our case, this would translate into ∼33,000 model runs, see Pianosi et al., 2016) Our choice was the result of a trade-off between maximizing the exploration of the parameter space and limiting computational time

(Pianosi et al., 2016). In order to limit the impact of this choice on our results, we sampled parameter values using the Latin-Hypercube method as implemented in SAFE (Pianosi et al., 2015), a quasi-random approach that repeats random sampling several times (in our case, five) and chooses as final sample the one that is more evenly distributed across the parameter space. Results reported here were qualitatively similar to a preliminary run performed using 5000 samples, evidence that our procedure had reached convergence.

5. Conclusions

Expanding Dynamic Identifiability Analysis to measure the information content of ground-based measurements of temperature, relative humidity, solar radiation, soil-water content, and SWE allowed us to identify optimal values for more parameter than relying on streamflow alone (seven instead of two). By identifying optimal values with more informative attributes, parameters also showed less temporal variability than those calibrated on streamflow. In particular dew-point-temperature measurements better constrained phase-partitioning parameters than using streamflow. Tuning parameters of a hydrologic-forecasting model based on distributed ground-based measurements allowed the model to better simulate key drivers of the surface water budget such as air temperature and precipitation-phase partitioning. However, intermediate processes in the water budget (evapotranspiration, snow accumulation and melt, radiation distribution, and sub-surface water storage) emerged as key sources of predictive uncertainty for distributed-parameter models that even multi-attribute Identifiability Analysis could not fully constrain. Integrating ground-based measurements with remote-sensing platforms and with improved, physics-based parametrizations is highly needed to overcome these sources of uncertainty and thus increase the overall predictive accuracy of distributed-parameter hydrologic models in mountain headwaters.

Acknowledgments

This work was supported by the California Energy Commission under contract EPC-14-067, Pacific Gas & Electric Co, the California Department of Water Resources, and the UC Water Security and Sustainability Research Initiative. Tessa Maurer was supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE 1106400. We would like to thank Kevin Richards (PG&E) for his continuous support and advice throughout this work. Analyses in this work benefited from the SAFE toolbox.

References

Abebe, N. A., Ogden, F. L., & Pradhan, N. R. (2010). Sensitivity and uncertainty analysis of the conceptual HBV rainfall-runoff model: Implications for parameter estimation. *Journal of Hydrology*, *389*, 301 – 310. doi:https://doi.org/10.1016/j.jhydrol.2010.06.007.

Andreadis, K. M., Clark, E. A., Lettenmaier, D. P., & Alsdorf, D. E. (2007). Prospects for river discharge and depth estimation through assimilation of swath-altimetry into a raster-based hydrodynamics model. *Geophysical Research Letters*, *34*. doi:10.1029/2007GL029721.

Andreadis, K. M., & Lettenmaier, D. P. (2006). Assimilating remotely sensed snow observations into a macroscale hydrology model. *Advances in Water Resources*, *29*, 872 – 886. doi:https://doi.org/10.1016/j.advwatres.2005.08.004.

Anghileri, D., Voisin, N., Castelletti, A., Pianosi, F., Nijssen, B., & Lettenmaier, D. P. (2016). Value of long-term streamflow forecasts to reservoir operations for water supply in snow-dominated river catchments. *Water Resources Research*, *52*, 4209–4225. doi:10.1002/2015WR017864.

Avanzi, F., De Michele, C., Ghezzi, A., Jommi, C., & Pepe, M. (2014). A processing-modeling routine to use SNOTEL hourly data in snowpack dynamic models. *Advances in Water Resources*, *73*, 16–29. doi:10.1016/j.advwatres.2014.06.011.

Avanzi, F., Maurer, T., Malek, S., Glaser, S. D., Bales, R. C., & Conklin, M. H. (2018). *Feather River Hydrologic Observatory: Improving Hydrological Snowpack Forecasting for Hydropower Generation Using Intelligent Information Systems*. Technical Report California's Fourth Climate Change Assessment, California Energy Commission.

Bales, R., Molotch, N. P., Painter, T. H., Dettinger, M. D., Rice, R., & Dozier, J. (2006). Mountain hydrology of the western United States. *Water Resources Research*, *42*, W08432. doi:10.1029/2005WR004387.

Bales, R. C., Goulden, M. L., Hunsaker, C. T., Conklin, M. H., Hartsough, P. C., O'Geen, A. T., Hopmans, J. W., & Safeeq, M. (2018). Mechanisms controlling the impact of multi-year drought on mountain hydrology. *Scientific Reports*, *8*, 690. doi:10.1038/s41598-017-19007-0.

Barnett, T. P., Adam, J. C., & Lettenmaier, D. P. (2005). Potential impacts of a warming climate on water availability in snow-dominated regions. *Nature*, *438*, 303–309. doi:10.1038/nature04141.

Barnhart, T. B., Molotch, N. P., Livneh, B., Harpold, A. A., Knowles, J. F., & Schneider, D. (2016). Snowmelt rate dictates streamflow. *Geophysical Research Letters*, *43*, 8006–8016. doi:10.1002/2016GL069690.

Bartolini, E., Allamano, P., Laio, F., & Claps, P. (2011). Runoff regime estimation at high-elevation sites: a parsimonious water balance approach. *Hydrology and Earth System Sciences*, *15*, 1661–1673. doi:10.5194/hess-15-1661-2011.

Bekele, E. G., & Nicklow, J. W. (2007). Multi-objective automatic calibration of SWAT using NSGA-II. *Journal of Hydrology*, *341*, 165 – 176. doi:https://doi.org/10.1016/j.jhydrol.2007.05.014.

Beven, K. (2001). How far can we go in distributed hydrological modelling? *Hydrology and Earth System Sciences*, *5*, 1–12. doi:10.5194/hess-5-1-2001.

Beven, K., & Freer, J. (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the glue methodology. *Journal of Hydrology*, *249*, 11 – 29. doi:https://doi.org/10.1016/S0022-1694(01)00421-8.

Beven, K., & Westerberg, I. (2011). On red herrings and real herrings: disinformation and information in hydrological inference. *Hydrological Processes*, *25*, 1676–1680. doi:10.1002/hyp.7963.

Beven, K. J. (2012). *Rainfall-Runoff Modelling: The Primer, 2nd Edition*. Wiley-Blackwell.

Blöschl, G., & Sivapalan, M. (1995). Scale issues in hydrological modelling: A review. *Hydrological Processes*, *9*, 251–290. doi:10.1002/hyp.3360090305.

Bongio, M., Avanzi, F., & De Michele, C. (2016). Hydroelectric power generation in an alpine basin: future water-energy scenarios in a run-of-the-river plant. *Advances in Water Resources*, *94*, 318 – 331. doi:https://doi.org/10.1016/j.advwatres.2016.05.017.

Boyle, D. P., Gupta, H. V., & Sorooshian, S. (2000). Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods. *Water Resources Research*, *36*, 3663–3674. doi:10.1029/2000WR900207.

Camici, S., Ciabatta, L., Massari, C., & Brocca, L. (2018). How reliable are satellite precipitation estimates for driving hydrological models: A verification study over the Mediterranean area. *Journal of Hydrology*, *563*, 950 – 961. doi:https://doi.org/10.1016/j.jhydrol.2018.06.067.

Clark, M. P., Slater, A. G., Barrett, A. P., Hay, L. E., McCabe, G. J., Rajagopalan, B., & Leavesley, G. H. (2006). Assimilation of snow covered area information into hydrologic and land-surface models. *Advances in Water Resources*, *29*, 1209 – 1221. doi:https://doi.org/10.1016/j.advwatres.2005.10.001.

Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous united states. *International Journal of Climatology*, *28*, 2031–2064. doi:10.1002/joc.1688.

De Michele, C., Avanzi, F., Ghezzi, A., & Jommi, C. (2013). Investigating the dynamics of bulk snow density in dry and wet conditions using a one-dimensional model. *The Cryosphere*, *7*, 433–444. doi:10.5194/tc-7-433-2013.

De Michele, C., & Ignaccolo, M. (2013). New perspectives on rainfall from a discrete view. *Hydrological Processes*, *27*, 2379–2382. doi:10.1002/hyp.9782.

Devineni, N., Sankarasubramanian, A., & Ghosh, S. (2008). Multimodel ensembles of streamflow forecasts: Role of predictor state in developing optimal combinations. *Water Resources Research*, *44*. doi:10.1029/2006WR005855.

Donovan, J., & Koczot, K. (2019). *User's manual for the Draper climate-distribution software suite with data-evaluation tools*. U.S. Geological Survey Techniques and Methods 7-C22, 55 p USGS.

Dressler, K. A., Leavesley, G. H., Bales, R. C., & Fassnacht, S. R. (2006). Evaluation of gridded snow water equivalent and satellite snow cover products for mountain basins in a hydrologic model. *Hydrological Processes*, *20*, 673–688. doi:10.1002/hyp.6130.

Fenicia, F., McDonnell, J. J., & Savenije, H. H. G. (2008). Learning from model improvement: On the contribution of complementary data to process understanding. *Water Resources Research*, *44*. doi:10.1029/2007WR006386.

Freeman, G. J. (1999). Runoff forecast error uncertainty and some ways it can affect snowmelt water scheduling decisions in the Sierra. In *Proceedings of the 67th Annual Western Snow Conference* (pp. 45–53).

Freeman, G. J. (2011). Climate Change and the Changing Water Balance for California's North Fork Feather River. In *Proceedings of the 79th Annual Western Snow Conference* (pp. 71–82).

Freeman, G. J. (2015). Planning Beyond California's Three-Year Drought - A 2015 Hydroelectric Planning Perspective. In *Proceedings of the 83rd Annual Western Snow Conference* (pp. 45–53).

Frei, C. (2014). Interpolation of temperature in a mountainous region using nonlinear profiles and non-euclidean distances. *International Journal of Climatology*, *34*, 1585–1605. doi:10.1002/joc.3786.

Froidurot, S., Zin, I., Hingray, B., & Gautheron, A. (2014). Sensitivity of precipitation phase over the swiss alps to different meteorological variables. *Journal of Hydrometeorology*, *15*, 685–696. doi:10.1175/JHM-D-13-073.1.

Garen, D. C. (1992). Improved Techniques in Regression-Based Streamflow Volume Forecasting. *Journal of Water Resources Planning and Management*, *118*, 654–670.

Gaudard, L., Avanzi, F., & Michele, C. D. (2018). Seasonal aspects of the energy-water nexus: The case of a run-of-the-river hydropower plant. *Applied Energy*, *210*, 604 – 612. doi:https://doi.org/10.1016/j.apenergy.2017.02.003.

Georgakakos, K. P., Graham, N. E., Carpenter, M., & Yao, H. (2004). Integrating climate-hydrology forecasts and multi-objective reservoir management for northern California. *Eos, Transactions American Geophysical Union*, *86*, 122–127. doi:10.1029/2005EO120002.

Ghasemizade, M., Baroni, G., Abbaspour, K., & Schirmer, M. (2017). Combined analysis of time-varying sensitivity and identifiability indices to diagnose the response of a complex environmental model. *Environmental Modelling & Software*, *88*, 22 – 34. doi:https://doi.org/10.1016/j.envsoft.2016.10.011.

Goulden, M. L., & Bales, R. C. (2014). Mountain runoff vulnerability to increased evapotranspiration with vegetation expansion. *Proceedings of the National Academy of Sciences*, *111*, 14071–14075. doi:10.1073/pnas.1319316111.

Gupta, H. V., Clark, M. P., Vrugt, J. A., Abramowitz, G., & Ye, M. (2012). Towards a comprehensive assessment of model structural adequacy. *Water Resources Research*, *48*. doi:10.1029/2011WR011044.

Gupta, H. V., Sorooshian, S., & Yapo, P. O. (1998). Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. *Water Resources Research*, *34*, 751–763. doi:10.1029/97WR03495.

Gupta, H. V., Sorooshian, S., & Yapo, P. O. (1999). Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. *Journal of Hydrologic Engineering*, *4*, 135–143. doi:10.1061/(ASCE)1084-0699(1999)4:2(135).

Harder, P., & Pomeroy, J. W. (2014). Hydrological model

uncertainty due to precipitation-phase partitioning methods. *Hydrological Processes*, *28*, 4311–4327. doi:10.1002/hyp.10214.

Harpold, A. A., Kaplan, M. L., Klos, P. Z., Link, T., McNamara, J. P., Rajagopal, S., Schumer, R., & Steele, C. M. (2017). Rain or snow: hydrologic processes, observations,prediction, and research needs. *Hydrology and Earth System Sciences*, *21*, 1–22. URL: https://www.hydrol-earth-syst-sci.net/21/1/2017/. doi:10.5194/hess-21-1-2017.

Harrison, B., & Bales, R. (2016). Skill Assessment of Water Supply Forecasts for Western Sierra Nevada Watersheds. *Journal of Hydrologic Engineering*, *21*, 04016002. doi:10.1061/(ASCE)HE.1943-5584.0001327.

Hatchett, B. J., Daudert, B., Garner, C. B., Oakley, N. S., Putnam, A. E., & White, A. B. (2017). Winter Snow Level Rise in the Northern Sierra Nevada from 2008 to 2017. *Water*, *9*, 899. doi:10.3390/w9110899.

Hay, L. E., Leavesley, G. H., Clark, M. P., Markstrom, S. L., Viger, R. J., & Umemoto, M. (2006). Step wise, multiple objective calibration of a hydrologic model for a snowmelt dominated basin. *JAWRA Journal of the American Water Resources Association*, *42*, 877–890. doi:10.1111/j.1752-1688.2006.tb04501.x.

Hoey, S. V., Nopens, I., van der Kwast, J., & Seuntjens, P. (2015). Dynamic identifiability analysis-based model structure evaluation considering rating curve uncertainty. *Journal of Hydrologic Engineering*, *20*, 04014072. doi:10.1061/(ASCE)HE.1943-5584.0000995.

Huang, G., Kadir, T., & Chung, F. (2012). Hydrological response to climate warming: The Upper Feather River Watershed. *Journal of Hydrology*, *426-427*, 138 – 150. doi:https://doi.org/10.1016/j.jhydrol.2012.01.034.

Jennings, K., Winchell, T. S., Livneh, B., & Molotch, N. P. (2018). Mspatial variation of the rain–snow temperature threshold across the northern hemisphere. *Nature Communications*, *9*. doi:10.1038/s41467-018-03629-7.

Johnson, J. B., & Marks, D. (2004). The detection and correction of snow water equivalent pressure sensor errors. *Hydrological Processes*, *18*, 3513–3525. doi:10.1002/hyp.5795.

Kirchner, J. W. (2006). Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. *Water Resources Research*, *42*. doi:10.1029/2005WR004362.

Klemeš, V. (1977). Value of information in reservoir optimization. *Water Resources Research*, *13*, 837–850. doi:10.1029/WR013i005p00837.

Kling, H., Fuchs, M., & Paulin, M. (2012). Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *Journal of Hydrology*, *424-425*, 264 – 277. doi:https://doi.org/10.1016/j.jhydrol.2012.01.011.

Knowles, N., Dettinger, M. D., & Cayan, D. R. (2006). Trends in Snowfall versus Rainfall in the Western United States. *Journal of Climate*, *19*, 4545–4559. doi:http://dx.doi.org/10.1175/JCLI3850.1.

Koczot, K. M., Jeton, A. E., McGurk, B., & Dettinger, M. D. (2004). *Precpitation-Runoff Processes in the Feather River Basin, Northeastern California, with Prospects for Streamflow Predictability, Water Years 1971-1997*. Scientific Investigations Report 5202 USGS.

Laiolo, P., Gabellani, S., Rebora, N., Rudari, R., Ferraris, L., Ratto, S., Stevenin, H., & Cauduro, M. (2014). Validation of the flood-proofs probabilistic forecasting system. *Hydrological Processes*, *28*, 3466–3481. doi:10.1002/hyp.9888.

Lawrence, M. G. (2005). The Relationship between Relative Humidity and the Dewpoint Temperature in Moist Air: A Simple Conversion and Applications. *Bulletin of the American Meteorological Society*, *86*, 225–234. doi:10.1175/BAMS-86-2-225.

Leavesley, G. H., Lichty, R. W., Troutman, B. M., & Saindon, L. G. (1983). *Precipitation-Runoff Modeling System: User's Manual*. Technical Report U.S. Geological Survey.

Lehning, M., Völksch, I., Gustafsson, D., Nguyen, T. A., Stähli, M., & Zappa, M. (2006). ALPINE3D: a detailed model of mountain surface processes and its application to snow hydrology. *Hydrological Processes*, *20*, 2111–2128. doi:10.1002/hyp.6204.

Li, H., Luo, L., Wood, E. F., & Schaake, J. (2009). The role of initial conditions and forcing uncertainties in seasonal hydrologic forecasting. *Journal of Geophysical Research: Atmospheres*, *114*. doi:10.1029/2008JD010969.

Malek, S. A., Avanzi, F., Brun-Laguna, K., Maurer, T., Oroza, C. A., Hartsough, P. C., Watteyne, T., & Glaser, S. D. (2017). Real-Time Alpine Measurement System Using Wireless Sensor Networks. *Sensors*, *17*. doi:10.3390/s17112583.

Margulis, S. A., Cortés, G., Girotto, M., & Durand, M. (2016). A Landsat-Era Sierra Nevada Snow Reanalysis (1985-2015). *Journal of Hydrometeorology*, *17*, 1203–1221. doi:10.1175/JHM-D-15-0177.1.

Marks, D., Winstral, A., Reba, M., Pomeroy, J., & Kumar, M. (2013). An evaluation of methods for determining during-storm precipitation phase and the rain/snow transition elevation at the surface in a mountain basin. *Advances in Water Resources*, *55*, 98 – 110. doi:https://doi.org/10.1016/j.advwatres.2012.11.012.

Markstrom, S. L., Hay, L. E., & Clark, M. P. (2016). Towards simplification of hydrologic modeling: identification of dominant processes. *Hydrology and Earth System Sciences*, *20*, 4655–4671. doi:10.5194/hess-20-4655-2016.

Markstrom, S. L., Regan, R. S., Hay, L. E., Viger, R. J., Webb, R. M., Payn, R. A., & LaFontaine, J. H. (2015). *PRMS-IV, the Precipitation-Runoff Modeling System, Version 4*. Technical Report U.S. Geological Survey Techniques and Methods. doi:10.3133/tm6B7.

Massari, C., Camici, S., Ciabatta, L., & Brocca, L. (2018). Exploiting Satellite-Based Surface Soil Moisture for Flood Forecasting in the Mediterranean Area: State Update Versus Rainfall Correction. *Remote Sensing*, *10*. doi:10.3390/rs10020292.

Mitchell, K. E., Lohmann, D., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., Cosgrove, B. A., Sheffield, J., Duan, Q., Luo, L., Higgins, R. W., Pinker, R. T., Tarpley, J. D., Lettenmaier, D. P., Marshall, C. H., Entin, J. K., Pan, M., Shi, W., Koren, V., Meng, J., Ramsay, B. H., & Bailey, A. A. (2004). The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. *Journal of Geophysical Research: Atmospheres*, *109*. doi:10.1029/2003JD003823.

Mote, P. W. (2003). Trends in snow water equivalent in the Pacific Northwest and their climatic causes. *Geophyical Research Letters*, *30*, 1601. doi:10.1029/2003GL017258.

Mote, P. W., Li, S., Lettenmaier, D. P., Xiao, M., & Engel, R. (2018). Dramatic declines in snowpack in the western US. *npj Climate and Atmospheric Science*, *1*. doi:10.

1038/s41612-018-0012-1.

Musselmann, K. N., Clark, M. P., Liu, C., Ikeda, K., & Rasmussen, R. (2017). Slower snowmelt in a warmer world. *Nature Climate Change*, *7*, 214 – 219. doi:10.1038/nclimate3225.

Nijzink, R. C., Almeida, S., Pechlivanidis, I. G., Capell, R., Gustafssons, D., Arheimer, B., Parajka, J., Freer, J., Han, D., Wagener, T., Nooijen, R. R. P., Savenije, H. H. G., & Hrachowitz, M. (2018). Constraining conceptual hydrological models with multiple information sources. *Water Resources Research*, . doi:10.1029/2017WR021895.

Oroza, C. A., Zheng, Z., Glaser, S. D., Tuia, D., & Bales, R. C. (2016). Optimizing Embedded Sensor Network Design for Catchment-scale Snow-depth Estimation Using LiDAR and Machine Learning. *Water Resources Research*, *52*, 8174–8189. doi:10.1002/2016WR018896.

Pagano, T., Garen, D., & Sorooshian, S. (2004). Evaluation of Official Western U.S. Seasonal Water Supply Outlooks, 1922-2002. *Journal of Hydrometeorology*, *5*, 896–909. doi:10.1175/1525-7541(2004)005<0896:EOOWUS>2.0.CO;2.

Pagano, T. C., Wood, A. W., Ramos, M.-H., Cloke, H. L., Pappenberger, F., Clark, M. P., Cranston, M., Kavetski, D., Mathevet, T., Sorooshian, S., & Verkade, J. S. (2014). Challenges of operational river forecasting. *Journal of Hydrometeorology*, *15*, 1692–1707. doi:10.1175/JHM-D-13-0188.1.

Parajka, J., Merz, R., & Blöschl, G. (2005). A comparison of regionalisation methods for catchment model parameters. *Hydrology and Earth System Sciences*, *9*, 157–171. doi:10.5194/hess-9-157-2005.

Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., & Wagener, T. (2016). Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environmental Modelling & Software*, *79*, 214 – 232. doi:https://doi.org/10.1016/j.envsoft.2016.02.008.

Pianosi, F., Sarrazin, F., & Wagener, T. (2015). A Matlab toolbox for Global Sensitivity Analysis. *Environmental Modelling & Software*, *70*, 80 – 85. doi:10.1016/j.envsoft.2015.04.009.

Pianosi, F., & Wagener, T. (2016). Understanding the time-varying importance of different uncertainty sources in hydrological modelling using global sensitivity analysis. *Hydrological Processes*, *30*, 3991–4003. doi:10.1002/hyp.10968.

Pool, S., Vis, M., & Seibert, J. (2018). Evaluating model performance: towards a non-parametric variant of the Kling-Gupta efficiency. *Hydrological Sciences Journal*, *63*, 1941–1953. doi:10.1080/02626667.2018.1552002.

Rajib, M. A., Merwade, V., & Yu, Z. (2016). Multi-objective calibration of a hydrologic model using spatially distributed remotely sensed/in-situ soil moisture. *Journal of Hydrology*, *536*, 192 – 207. doi:https://doi.org/10.1016/j.jhydrol.2016.02.037.

Renard, B., Kavetski, D., Kuczera, G., Thyer, M., & Franks, S. W. (2010). Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. *Water Resources Research*, *46*. doi:10.1029/2009WR008328.

Roche, J. W., Bales, R. C., Rice, R., & Marks, D. G. (2018). Management Implications of Snowpack Sensitivity to Temperature and Atmospheric Moisture Changes in Yosemite National Park, CA. *JAWRA Journal of the American Water Resources Association*, . doi:10.1111/1752-1688.12647.

Rolland, C. (2003). Spatial and Seasonal Variations of Air Temperature Lapse Rates in Alpine Regions. *Journal of Climate*, *16*, 1032–1046. doi:10.1175/1520-0442(2003)016<1032:SASVOA>2.0.CO;2.

Rosenberg, E. A., Wood, A. W., & Steinemann, A. C. (2011). Statistical applications of physically based hydrologic models to seasonal streamflow forecasts. *Water Resources Research*, *47*. doi:10.1029/2010WR010101.

Rössler, O., Froidevaux, P., Börst, U., Rickli, R., Martius, O., & Weingartner, R. (2014). Retrospective analysis of a nonforecasted rain-on-snow flood in the Alps - a matter of model limitations or unpredictable nature? *Hydrology and Earth System Sciences*, *18*, 2265–2285. doi:10.5194/hess-18-2265-2014.

Saksa, P. C., Conklin, M. H., Battles, J. J., Tague, C. L., & Bales, R. C. (2017). Forest thinning impacts on the water balance of Sierra Nevada mixed-conifer headwater basins. *Water Resources Research*, *53*, 5364–5381. doi:10.1002/2016WR019240.

Samaniego, L., Kumar, R., & Attinger, S. (2010). Multi-scale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resources Research*, *46*. doi:10.1029/2008WR007327.

Santos, L., Thirel, G., & Perrin, C. (2018). Technical note: Pitfalls in using log-transformed flows within the KGE criterion. *Hydrology and Earth System Sciences*, *22*, 4583–4591. doi:10.5194/hess-22-4583-2018.

Schaefli, B., & Gupta, H. (2007). Do Nash values have value? *Hydrological Processes*, *21*, 2075?–2080. doi:10.1002/hyp.6825.

Schaefli, B., Hingray, B., Niggli, M., & Musy, A. (2005). A conceptual glacio-hydrological model for high mountainous catchments. *Hydrology and Earth System Sciences*, *9*, 95–109. URL: http://www.hydrol-earth-syst-sci.net/9/95/2005/. doi:10.5194/hess-9-95-2005.

Schaefli, B., Nicótina, L., Imfeld, C., Da Ronco, P., Bertuzzo, E., & Rinaldo, A. (2014). SEHR-ECHO v1.0: a Spatially Explicit Hydrologic Response model for ecohydrologic applications. *Geoscientific Model Development*, *7*, 2733–2746. URL: https://www.geosci-model-dev.net/7/2733/2014/. doi:10.5194/gmd-7-2733-2014.

Serreze, M. C., Clark, M. P., Armstrong, R. L., McGinnis, D. A., & Pulwarty, R. S. (1999). Characteristics of the western United States snowpack from snowpack telemetry (SNOTEL) data. *Water Resources Research*, *35*, 2145–2160. doi:10.1029/1999WR900090.

Silvestro, F., Gabellani, S., Rudari, R., Delogu, F., Laiolo, P., & Boni, G. (2015). Uncertainty reduction and parameter estimation of a distributed hydrological model with ground and remote-sensing data. *Hydrology and Earth System Sciences*, *19*, 1727–1751. doi:10.5194/hess-19-1727-2015.

Silvestro, F., Parodi, A., Campo, L., & Ferraris, L. (2018). Analysis of the streamflow extremes and long-term water balance in the liguria region of italy using a cloud-permitting grid spacing reanalysis dataset. *Hydrology and Earth System Sciences*, *22*, 5403–5426. doi:10.5194/hess-22-5403-2018.

Tang, Q., & Lettenmaier, D. P. (2010). Use of satellite snow-cover data for streamflow prediction in the Feather River Basin, California. *International Journal of Remote Sensing*, *31*, 3745–3762. doi:10.1080/01431161.2010.483493.

Vrugt, J. A., Bouten, W., Gupta, H. V., & Sorooshian, S. (2002). Toward improved identifiability of hydrologic

model parameters: The information content of experimental data. *Water Resources Research*, *38*, 48–1–48–13. doi:10.1029/2001WR001118.

Wagener, T., & Gupta, H. V. (2005). Model identification for hydrological forecasting under uncertainty. *Stochastic Environmental Research and Risk Assessment*, *19*, 378–387. doi:10.1007/s00477-005-0006-5.

Wagener, T., McIntyre, N., Lees, M. J., Wheater, H. S., & Gupta, H. V. (2003). Towards reduced uncertainty in conceptual rainfall-runoff modelling: dynamic identifiability analysis. *Hydrological Processes*, *17*, 455–476. doi:10.1002/hyp.1135.

Wanders, N., Bierkens, M. F. P., de Jong, S. M., de Roo, A., & Karssenberg, D. (2014). The benefits of using remotely sensed soil moisture in parameter identification of large-scale hydrological models. *Water Resources Research*, *50*, 6874–6891. doi:10.1002/2013WR014639.

Wi, S., Yang, Y. C. E., Steinschneider, S., Khalil, A., & Brown, C. M. (2015). Calibration approaches for distributed hydrologic models in poorly gaged basins: implication for streamflow projections under climate change. *Hydrology and Earth System Sciences*, *19*, 857–876. doi:10.5194/hess-19-857-2015.

Wood, A. W., & Schaake, J. C. (2008). Correcting errors in streamflow forecast ensemble mean and spread. *Journal of Hydrometeorology*, *9*, 132–148. doi:10.1175/2007JHM862.1.

Yapo, P. O., Gupta, H. V., & Sorooshian, S. (1998). Multi-objective global optimization for hydrologic models. *Journal of Hydrology*, *204*, 83 – 97. doi:https://doi.org/10.1016/S0022-1694(97)00107-8.

Zhang, R., Liu, J., Gao, H., & Mao, G. (2018). Can multi-objective calibration of streamflow guarantee better hydrological model accuracy? *Journal of Hydroinformatics*, . doi:10.2166/hydro.2018.131.

Zhang, Z., Glaser, S., Bales, R., Conklin, M., Rice, R., & Marks, D. (2017). Insights into mountain precipitation and snowpack from a basin-scale wireless-sensor network. *Water Resources Research*, *53*, 6626–6641. doi:10.1002/2016WR018825.